GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 19, 2003, 15:29:20 ; Search time 3325 Seconds
(without alignments)
5533.380 Million cell updates/sec

Title: US-09-494-297-2
Perfect score: 3945
Sequence: 1 MKKTRFPNKLNTLNTQRVLS.........IAGISLGIWGIHITIRIRKHD 757

Scoring table:
BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop   6.0 ; Delext   7.0

Searched: 22781392 seqs, 1215223856 residues

Total number of hits satisfying chosen parameters: 4556284

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/Cgn2_1/USPTO_spool/US09494297/runat_13082003_122947_28105/app_query.fasta_1.903
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09494297.@CGN_1_1.2743.@runat_13082003.122947_28105 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: em_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_estfun:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_pro:*
26: em_gss_rod:*
27: em_gss_phg:*
28: em_gss_vrl:*
27: em_gss_phg:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	128	3.2	3344	11	BC028681	Homo sapi
2	123.5	3.1	3778	11	AK079441	Mus muscu
3	121	3.1	3617	11	AK084780	Mus muscu
4	120	3.0	2138	11	AK031936	Mus muscu
5	120	3.0	3959	11	AK053071	Mus muscu
6	118.5	3.0	4594	11	AK036592	Mus muscu
7	116	3.0	704	12	BM169720	EST572243
8	115.5	2.9	708	12	BM674993	UI-E-EJO-
9	114.5	2.9	863	28	AZ681177	ENTMR45TR
10	112.5	2.9	945	13	BX350683	Homo sapi
11	112	2.8	2421	11	BC008707	Homo sapi
12	112	2.8	2430	11	BC015953	Homo sapi
13	111.5	2.8	1493	11	AK076636	Mus muscu
14	111.5	2.8	2167	11	AK077916	Mus muscu
15	111	2.8	889	28	BH132924	ENTMX37TF
16	111	2.8	3250	11	BC047878	Homo sapi
17	110	2.8	2419	11	AK035141	Mus muscu
18	109	2.8	2673	11	BC040379	Homo sapi
19	109	2.8	2677	11	AK016632	Homo sapi
20	109	2.8	3044	11	AK023358	Mus muscu
21	109	2.8	3318	11	AK032570	Mus muscu
22	109	2.8	3493	11	AK080043	Mus muscu
23	108.5	2.8	1085	12	BM913991	AGENCOURT
24	108.5	2.8	626	12	BE919552	EST423321
25	108	2.7	961	13	BX461833	BX461833
26	107.5	2.7	584	9	AV603854	AV603854
27	107.5	2.7	874	14	CD302015	AGENCOURT
28	106.5	2.7	1320	11	AK016441	Mus muscu
29	106.5	2.7	2955	11	AK089817	Mus muscu
30	106.5	2.7	4415	11	AK076369	Mus muscu
31	106	2.7	964	13	BX350916	Homo sapi
32	106	2.7	3046	11	AK046668	Mus muscu
33	106	2.7	3640	11	AK028842	Mus muscu
34	106	2.7	596	13	BQ591910	Homo sapi
35	105.5	2.7	783	10	BG661779	602679706
36	105.5	2.7	1557	28	BH770729	BH770729
37	105.5	2.7	1743	11	AY067850	Schmidtea
38	105	2.7	961	29	CNS06VHB	CNS06VHB
39	105	2.7	2880	11	BH770980	T3 end of
40	105	2.7	3416	11	AK041657	Mus muscu
41	104.5	2.6	686	13	BW211560	Mus muscu
42	104.5	2.6	741	13	BW212561	Mus muscu
43	104.5	2.6	748	13	BW209345	Mus muscu
44	104.5	2.6	2595	11	AK042706	Mus muscu
45	104	2.6	2613	11	AK028919	Mus muscu

ALIGNMENTS

RESULT 1
BC028681
LOCUS   BC028681   3344 bp   mRNA   linear   HTC 01-MAY-2002
DEFINITION Homo sapiens, retinoic acid induced 14, clone IMAGE:4822221, mRNA.
ACCESSION   BC028681
VERSION   BC028681.1 GI:20379528
KEYWORDS   HTC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3344)
AUTHORS   Strausberg,R.

```
TITLE      Direct Submission
JOURNAL    Submitted (29-APR-2002) National Institutes of Health, Mammalian
           Gene Collection (MGC), Cancer Genomics Office, National Cancer
           Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
           USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
           cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
           Toshiyuki and Piero Carninci (RIKEN)
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Institute for Systems Biology
           http://www.systemsbiology.org
           contact: amdan@systemsbiology.org
           Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
           Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
           This clone has the following problem: frame shifted.
           Clone distribution: MGC clone distribution information can be found
           through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
           Series: IRAK Plate: 33 Row: C Column: 23
           This clone was selected for full length sequencing because it
           passed the following selection criteria: matched mRNA gi: 13470085
FEATURES   Location/Qualifiers
source     1..3344
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="LocusID:26064"
           /db_xref="taxon:9606"
           /clone="IMAGE:4822221"
           /tissue_type="Testis"
           /clone_lib="NIH_MGC_97"
           /lab_host="DH10B"
           /note="Vector: pBluescript"
BASE COUNT      691 a    782 g    701 t
ORIGIN
        1170 a                              3344
```

```
US-09-494-297-2 (1-757) x BC028681 (1-3344)

Alignment Scores:
Pred. No.:                              11
Score:             0.0252       Length:           3344
                   128.00       Matches:           151
Percent Similarity: 35.72%      Conservative:       98
Best Local Similarity: 21.66%   Mismatches:        254
Query Match:        3.24%       Indels:            195
DB:                 11          Gaps:               36
```

```
Qy    124 LysLysTrp---TyrLysLysHisAspGlyIleSerThrLysPheGluAspTyr------     140
          ||||||||||          ||||||          ||||||
Db   1156 AAGAAGTGGAACTCCAAAAAAAACGGACAAAGCCCAAGAGAAGCTTGACTCATACCATT     1215

Qy    141 -----AlaMet-SerProArg---IleThrGlyAspGluLeuAsnGlnLysLeuArgAl      157
               :||  ||||||      ||||||||  ||||
Db   1216 GAGTCGATGTCTCTCCCAAGATCAATAACTTCGACTCCACTATCGGGAAAG--GAATC     1272

Qy    157 aValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyIleMetGluGlyLeuGluProLe      177
          :|||          ||||||||          ||||
Db   1273 GGTATTTTTTGCTGAACCACCCTTCAAGGCTGAG------------------AT      1308
```

```
Qy    177 uAsnAlaIleArgValThrGlnGluAlaValTrpTyrTyrSerAspAsnAlaProIleSe      197
          :||||||                    ||||||              ||
Db   1309 CAGTTCTATACGAGAAAAACACAAGAGACAGACTA-----------AGTGACAGTACT     1353

Qy    197 rAsnProAspGluSerPheLysArgGluSer-------AsnLeuValSerTh      214
          ||  ||||||  ||||||          ||||:::||
Db   1354 AGGTGCTGATAGCTTATTGGATATAAGTCTGAAGCTGACCAACAAGATCTTCTCTCT     1413

Qy    214 rSerGlnLeuSerLeuMetArgGlnAlaLeu-----LysGlnLeuIleLeuAspProAsnLe      232
          ||          |||:|||:||:|||||          |||
Db   1414 ATTGCAAGCAAAAGTTGCTTCCCTTACCTTACACAATAAGGAGTTACAGAT--AAATT     1470
```

```
Qy    232 uAlaThrLysMetProLysGlnValProAspAspPheGluGlnLeuSerIlePheGluSer-- 251
          |||||||              |||
Db   1471 ACAGGCCCAAATCACCAAGGAGGCGGAAGCAGAAGACCTAAGCTTTGACTCATACCAC 1530
```

```
Qy    252 ----GluAspLysGlyAspLysProThrThrAsnLysGlyGlyTyrGlnLeuAsnLeuSerGlyGlyIleGlyLysGlyIleAsnLeuLeuSerGlyGlyGlyLysGlyGlyGly 270
              ||||||||||||       |||
Db   1531 CCAAACTGACTTGGGCCCATCCCTGGGAAA----------------------------------- 1561

Qy    270 uValProThrLysProProThrProGlyAspProProMetProProAsnGlnProGlnThr 290
          ||||||||||          ||          ::
Db   1562 ----CCTGGTGAA---ACCTCTCCCCAGACTCCAAATCATC 1596

Qy    290 rThrSerValLeuLeuArgLysTyrArgLeuAlaIleGlyAspTyrSerGlyLysLeuLeuGlyIyal 310
          ||||||          |||
Db   1597 TCCATCTGTCTTAATA----CATTCTTTAGGT---------------------------- 1624

Qy    310 aThrLeuGlnLeuThrGlyLysAspAsnValAsnSerPheGlnAlaArgValPheSerSerAs 330
          ::|          |||
Db   1625 --------AAATCCACTACTGACAATGATGTCAGAATTCAGCAACTGACAAGAGAATTTTCCA 1677

Qy    330 nAspIleGlyLeuArgIleGluLeuSerAsp-------------------------- 340
          ||:::          |||:::
Db   1678 AGATCTACAGAGAGAGATTAGAGAGCTCTGAACCAGGAAAACAGCTACAGGTCGAACT 1737
```

```
Qy    341 --------------------GlyThrTyrThrLeuThrGluLeuAsnse 350
                              |||:|
Db   1738 CCAATCCCGAAGGGCCAGAACGTGTATGCTAAACAACACTGGAGTTCAGAGAACAGCTC 1797

Qy    350 rProAlaGlyTyrTyrSerIleAlaGluProIleThrPheLysValGluAlaGlyValTy 370
          |          |||
Db   1798 TGACCTCAGCCACGAAAACTTAAAGAAACTTAGAGAGCAAATACAGAGAGGCTAGAAGAGT 1857

Qy    370 rThrIleLeuAspGlyLysGlnIleGluGluAsnProAsnLysLysGlyIleValGluProTyrSe 390
          :::          |||          |||
Db   1858 CCTTAGTGTGCAG---AAGCAGATGAAA-------------CTCGGTCTTGTCTCACCTGAAG 1905

Qy    390 rValGluAlaTyrAsnAspProPheGlnPheSerValLeuThrThrGlnAsnTyrAlaLy 410
          |||          |||
Db   1906 CATGGGAATAATTATTCACATTTCACCGACCTGAGGGTCACGGAAGAGGA-AATAAATGTGC 1964
```

```
Qy    410 sPheTyrTyrAlaLysAsnLysGlySerGlnValValTyrCysPheAsnAlaAs 430
          |          |||
Db   1965 T-----------------------------GGTAGAGAGGGAGAAAGGTACAGTGATTA 2041

Qy    430 pLeuLysSerProProAspSerGlnAsp---GlyGlyLysThrMetProAspPheTh 449
          |          |||||||
Db   1998 GTGAAAGAAAATAAAAGAGAAGAGTTAGAGAGAAAAACT----------- 2040

Qy    449 rThrGlyGluValAllysTyrThrHisIleAlaGlyArgAspLeuPheLysTyrThr-Vall 469
          |||||:::          |||
Db   2041 -------------------------------GGTAGAGAGGGAGAAAGGTACAGTGATTA 2069
```

```
Qy    469 ySProArgAspThrAspProThrPheLeuLysHisIleLysIleLysLysValIleGluLysG 489
          |||||          |||:::
Db   2070 AGCCACCTGTGGGAGAGTACGAGAAATGAAGTTCAATATTGCTCTGTTATTGAGAATTA 2129

Qy    489 lyTyrArgGlyLysGlyGlyGlnAlaIle------------GluTyrSerG 501
          ::|:|||||||          |||
Db   2130 TGAATAAGGAGAAAGCATTTTTGTTTGAGAAATTACCAAGAAGAAAAATCATGA 2189

Qy    501 lyLeuThrGluThrGlnLeuArgAlaAlaThrGlnLeuAlaIleGluTyrArgPheThrAspS 521
          |||          |||:::
Db   2190 AATTAAAAGACACACTAAAAAAGTCAGATGACACAGGAAGCC---------- 2237
```

```
Qy    521 erAla------------------GluLeuAspLysAsp--- 527
          |||||:          |||||:::|||
Db   2238 AAGCTGAGGACATGAAAGAAGGCCATGAAATAGGAGTGA 2297

Qy    528 -------------------LysLeuLysAspTyrHisGlyPheG 536
          :::|||:::|||||||
Db   2298 GCGAGCTGTCACAGCTGTACAAAGAAGACCCAGGCTGAGCTGAGGATTACAGGAAGAGGA 2357

Qy    536 LysAspMetAsnAspSerThrLeuAlaValAlaLysIleLeuValGluTyr------- 553
          :::          |||
Db      1                                                         -A  1
```

```
Db  2358 AATCTCTAGAGAGATGTCACA------------------------GCTGAATATATCCATAAAG 2396
Qy   553 laGlnAspSerAsnProProGlnLeuThrAspLeu---------------------AspPhe 567
                                                                   |||
Qy  2397 CAGAGCATGAGAAACTGAATGCAATTGACAACAGTGTCCAGGCTAAAGCAGAGATGCAC 2456
Db   567 heIleProAsnAsnLysTyrTyrGlnSerLeuIleGlyThrGlnTrpHisProGluAspL 587
         ::::|:|||||
Qy  2457 TGTCTCGAAATGAAATGCTCAGTATTCAAAAAGTGTTGAATGAGTTGACCCAGCTCAAACAAC 2516
Db   587 euValAspIleIleArgMetGluAspLysLysGluValIleProValThrHisAsnLeuT 607
          ||||||| :::  |||       ::: |||  ::|||
Qy  2517 TGGTTGGAT-------------CCACAAAAAGAGAAACTCTGTCTCTATCACACAAACATTTGC 2564
Db   607 hrLeuArgLysThrValThrGlyLeuAlaGlyArgThrLysAspPheHisProGluI 627
          :::|||     |||
Qy  2565 AAGTGATAACCACGCTG-------------CGGACTCCAGCAAAAGAATGGAAG 2606
Db   627 leGluLeuLysAsnAsnLysGlnGlnLeuLeuLeuSerGlnThrValLysThrAspLysThrA 647
          |||      |||  ::: |||        ::|||
Qy  2607 AAAAAAAATAAGCAATCTTAAGGAACACCTTGCAAGGAAGTAGCAACACATTGG 2666
Db   647 snLeuGluGluPheLysAspGlyLysAlaThrThrIleAsnLeuLysHisGlyGluSerLeuThrL 667
          :::       |||     :::|||       ::: :::|||
Qy  2667 AGAAACAAACTCTTAGAAGAGAAAAAGCT-------------GCTATGACTG 2702
Db   667 euGlnGlyLeuPro------------GluGlyTyrSerTyrLeuValLysGluThrAspGluGlyT 686
          ||| |||     :::||| |||         :::||||||
Qy  2703 ATGCAATGGTACCTCGGTCTTCCTATCAAAAACTCCAGTCATCCTTAGAGAGTGAA----2758
Db   686 yrLysValLysValAsnSerGlnLeuValAlaIleAlaThrValSerLysTyrLysThrGlyIleT 706
                                        |||        |||
Qy  2759 ------------GTGAGTGTGTTGGCATCGAAATTAAAGGAATCTGTGAAAAGAGAGAAGGTCC 2813
Db   706 hrSerAsp------------GluThrLeuAlaPheGluAsnAsnLysGluPro 720
          |||     :::||| :::            |||||| :
Qy  2814 ATTCAGAGTTGTCCAGATTAGAAGTGAAGAGAGAAAAGAAAAAAGAAAAATA 2873
Db   720 al------------ValProThrGlyValAspGlnLysIleIleAsnGlyTyrLeu 733
          ::         :::::|||           |||
Qy  2874 TTCAGACTCTCTTGAAATCCAAAGCAAGAGAAGTAAAAATGAACTTCTG 2920
```

RESULT 2

```
LOCUS       AK079441                3778 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult female vagina cDNA, RIKEN full-length enriched
            library, clone:9930026E20 product:desmocollin 1, full insert
            sequence.
ACCESSION   AK079441
VERSION     AK079441.1  GI:26098508
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   MEDLINE   99279253
   PUBMED    10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   MEDLINE   20490374
   PUBMED    11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   MEDLINE   20530913
   PUBMED    11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
   MEDLINE   21085660
   PUBMED    11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3778)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,W.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kamiya,I., Kasukawa,T.,
            Koya,S., Kurihara,C., Kawai,J., Kojima,Y., Kondo,S., Kono,H., Kouda,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,Y., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3778
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
```

```
                    /db_xref="FANTOM_DB:9930026E20"
                    /db_xref="taxon:10090"
                    /clone="9930026E20"
                    /sex="female"
                    /tissue_type="vagina"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
misc_feature        1..3778
                    /note="desmocollin 1 (MGD|MGI:109173, GB|NM_013504,
                    evidence: BLASTN, 99%, match=2955)"

BASE COUNT   1148 a   829 c   811 g   989 t   1 others
ORIGIN

Alignment Scores:
Pred. No.:              0.0971
Score:                  123.50          Length:        3778
Percent Similarity:     31.62%          Matches:        144
Best Local Similarity:  19.46%          Conservative:    90
Query Match:            3.138           Mismatches:     263
DB:                     11              Indels:         243
                                        Gaps:            36

US-09-494-297-2 (1-757) x AK079441 (1-3778)
```

```
Qy   102  ArgSerTyrGlnValTyrCysPheAsnLeuLysLysAlaPheProLeuGlySerAsp---  120
              ||||||   :::.|||||||||||||  ||||||||||  ||||||.||::|
Db   532  AGAAGCTAGAAGTTGTACGTCACGAAGAGAAGAAAAAAGGTTT---TTAGGAAAGACATA  588

Qy   121  ------------SerSerValLysLysTrpTyrLysLysHisAspGlyIle            133
                        :|||||||||||||||||||||||||||||||
Db   589  CCAAAGAGCCAGTACACAACCGCCAGTAAGAAGAAGAGATGGGCTCCT---            636

Qy   134  SerThrLysPheGluAspTyrAlaMetSerProArgIleThrGlyAspGluLeuAsnGln   153
              ::.:|||||   :::|
Db   637  CCATGTTCACTGATGGGAGGAACTCTTTGGGTCCA---                        669

Qy   154  LysLeuArgAlaValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyIleMetGluGly   173
              |||
Db   670  ---TTTCACAACACAATTCAGCAGATCCAATCTGAT                          702

Qy   174  LeuGluProLeuAsnAlaIleAlaIleArgValThrGlnGluAlaValThrProTyrTyrSerAspAsn  193
              :::|||||||||
Db   703  ---GCTGCCCAGAATTACACCATACTTTTACTCTATCAGT                      738

Qy   194  AlaPro------IleSerAsnProAspGluSerPheLysArgGluSerGluSerAsnLeu   211
              |||         :::.:|
Db   739  GGACCAGGAGTAGACACCAAGAGCCTTACAATTGTTCTACATAGAGAAAGACACCTGAC   798

Qy   212  ValSerThrSerGlnLeuSerLeuMetArgGlnAlaLeuLeuLysGlnLeuIleLeuAspProAsn  231
              :::
Db   799  ATCTATTTGTACCCGA---AGCATTGACCGTGAACAGTATGACCAATTTTGGTATATGGA   855

Qy   232  LeuAlaThrLysMetProLysGlyValProAspAspPheGlnLeuSerIlePheGluSer   251
              |||||         :::.:||
Db   856  TATGCCAACAACTGCTGATGGCTATGCCCCAGACTACCCTCTCCCCTGCTGTTAAGGTT   915

Qy   252  GluAspLysGlyAspLysGlyAspLysGlyTyrAsnLysGlyTyr                 262
              ||||||      |||
Db   916  GAAGATGACAATGAC---AATGCCCCATATTTTGAAACCAAAATTAACAGTTTTTAGT    969

Qy   263  ------GlnAsnLeuLeuSerGlyLysGlyLeuValProThrLysProProThrProGlyTyrAsp  280
              ::.:||||
Db   970  GTGCCTGAAAATTCCGACTCTGAAACTTCAGTGGACAGTGCCATAGACAAGGAT        1029

Qy   281  ProPro------MetProProAsnGln                                   287
              |||         :::.:|||
Db  1030  GAGCCAGGAGTAGACACTCTACACACTCGTCTGGAAGTAGACAAAATCTTACAACAACTCCCAGACCAG 1089

Qy   288  ProGlnThrThrSerValLeuIleLeuArgLysTyrAlaIleGlyAspTyrSerLysLeuLeu 307
              |||
Db  1090  CCA---                                                        1128

Qy   308  GluGlyAlaThrLeuGlnLeuThrLeuTyrAspAsnValAsnSerProGlnAlaArgValPhe  327
              |||:::.:||
Db  1128  AAGCACTTCTCCATCCACCGACACAGGTGTGATC                            1128
```

```
Db  1129  ACTACCACCACCTTTACTGGACGGAGAGAAATGTGACACTTACAAGCTA---GTGATG    1185

Qy   328  SerSerAsnAspIleGlyGluArgIle------GluLeuSerAspIleGlyThrTyrThr---  344
              ||||| :::   |||         :::.:||
Db  1186  GAGGTACGAGACAATGGGGGCCAGCCCTTTGGTCTCTGTTCCACACAGGAAATCACATC    1245

Qy   345  ---LeuThrGluLeuAsnSerProAlaGlyTyrSerIleAlaGlyProIleThrPheLys   363
              |||   :::.:||
Db  1246  TCACTGGAGGTGAGAATGACAACTCGCCATACTTCACCAAACTTCTTATACTACAGAA    1305

Qy   364  ValGlyLeuAlaGlyLys---                                          368
              ||||||   :::|
Db  1306  GTAGAGAGAAACACAGAAATTGATGTTGAGATTCTGACGACCAAGATTTG            1365

Qy   369  ------ValTyrThrIleLeuGluAspGlyLysGlnIleLeuGluAsn               380
              :::.:||
Db  1366  CCCAACACCCCCACTCGAAGCGGTTATACACAATCCTAAAGGGAACGAAAAACGGGAAC    1425

Qy   381  ---ProAsnLysGlu------                   IleValGluProTyr        389
              |||            ||||: :::||.|
Db  1426  TTCAAAATCACAACGGATCCAAATACAAATGAAGGGGGTCTTTGGTGTTGTCAAGCCACTG  1485

Qy   390  SerValGlnAlaTyrAsnAspPheGlu---GluPheSerValLeuThrThrGlnAsnTyr   408
              ||::.:                       :::.:||
Db  1486  AACTATGAAGTCCAGCGGCCGAAGTCACTCTGAAATTGGTGTCCTT---AATGAA        1536

Qy   409  AlaLysPheTyrTyrValTyrAlaLysAsnLysAsnGlySerGlnValValTyrCysPheAsn  428
              |||:::.:|||||
Db  1537  GCCCAGTTCACTAAGCCGTGTCCAACGCACCACCAACAATGCACCAACACTGTCACT      1596

Qy   429  AlaAspLeuLysSerProGluAspSerGlyAspGlyLysThrMetThrProAspPhe     448
              |||||:::||||      :::
Db  1597  GTGAAAATTAAA------GACAGAGATGAGGGCCCTGAGTGCCAGTC               1644

Qy   449  ThrThrGlyGluValLysTyrThrHisIleAlaGluGlyArgAspLeuPheLysTyrThrVal  468
              |||||:::.:||.|||||
Db  1645  AAAGTCATTCAGAGCAAAGATGGACTGCCACGGGGCCAAGAGACTCTGGGTAC---      1698

Qy   469  LysProArgAspThrAspProAspThrPheLeuGluLysHisIleLysLysValIleGluLys  488
              :::
Db  1699  ------AAAGCCGTGGACCCACAGACA---                                1719

Qy   489  GlyTyrArgGluLysGlyGlnAlaLeuAlaIleGluGlyTyrSerGlyLeuThrGlnLeuArg  508
              |||:::
Db  1720  ---AGCAGTGGTGAAGGCTTAAGGTAT---                                1743

Qy   509  AlaAlaThrArgGlnLeuAlaIleGluTyrTyrPheThrAspSerAlaGluLeuAspLysAspLys  528
              :::
Db  1744  ---GAGATG                                                     1749

Qy   529  LeuLysIleAspTyrHisGlyPheGlyAspMetAsnAspSerThr-----LeuAlaValAla  546
              |||||:::.:|||                              :::
Db  1750  GTAGGAGATGAGAGTAAATTGGTTTGAGATTAACAAAATCACCGGAGACACTGTG       1809

Qy   547  LysIleLeuValGluTyrAlaGlnAsnProProGlnLeuThrAspLeuAspPhe        566
              |||||:::                                :::
Db  1810  AAAGTACTGGACAGAGAGTCGAAG---                                   1833

Qy   567  PheIleProAsnAsnLysTyrGlnSerLeuIleGluGlyThrGlnIleTrpHisProGluAsp  586
              |||||:::.:|||||
Db  1834  TTTGTGAAAAACAACAACAGTACAAATATTTCAGTAGTGCGACAGATACAGCTGGCCGGATCT 1893

Qy   587  LeuValAsp------IleIleArgMetGluAspLysLysGluValIleIleProValThrHis  604
              ||||:::        ::.:|||
Db  1894  TGCACAGGAACACTGGTAGTTCTTCGGAAGATTTTAAATGAC---                 1938

Qy   605  AsnLeuThrLeuArgLysThrValThrGlyLeuAlaGlyAspArgThrLysAspPhe---  623
              |||.|:::.:|||||                                     :::
Db  1939  CCACCACAGATTGACAAGGAAGAGAGTGACC------ATTTGTCAGCAGGAAGATTTTGCT  1992

Qy   624  ---HisPheGluIleGluLeu                                          629
              |||:::.:|||
Db  1993  GTTTTGGAACCTATAGATTTAGATGGGCCCAGATAAARGGTGTCCACCTTTTCAGTTCCTCTG 2052
```

```
Qy   630 LysAsnAsnLysGlnGluLeuLeuSerGlnThrValLysThrAspLysThrAsnLeuGlu 649
             ||||:||
Db  2053 GATAATTCTTCCAGCAAACTTTGGACT------------------------CTAGAA 2085

Qy   650 PheLysAspGlyLysAlaThrIleAsnLeuLysHisGlyGluSerLeuThrLeuGlnGly 669
             :::||||||||||||||...||||
Db  2086 TCACAGGATGGTAAACGTGCCATT---CTTCGTCAACGACGACAATCTTAATTAT----- 2136

Qy   670 LeuProGlyGluGlyTyrSerTyrLeuValLysGluThrHrAspSerGluGlyTyrLysVallys 689
             ------||||||||...
Db  2137 ------AACTATTATTCTGTGCCAAATCCAAATACAGACAGCATGGTTTTTCCGCAAAA 2190

Qy   690 ---ValAsnSerGlnGluValAlaAlaAsnAlaThrValSerLysThrGlyIleThrSerAsp 708
             |||  |||
Db  2191 CATGTATTATCAGTGCGAGTGTGTGTGACTGTACAACT------------CCAACAGAA 2235

Qy   709 GluThrLeuAlaPheGluAsnAsnLysGluProValValProThrGlyValAspGlnLys 728
             :::||| |||  :::
Db  2236 TCTAGAATGGCTGTTAAGGAAGAGAGG----------------------GATGCTAAA 2271

Qy   729 IleAsnGlyTyrLeuAlaLeuIleValIleLeuSerLeuGlyIleLeuTrpGlyIle 748
             |||
Db  2272 CCAAAT---AATAATCCTTGGGAAATGGGCCATT 2301
```

RESULT 3

```
LOCUS       AK084780                3617 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched
            library, clone:D330041I19 product:TITIN homolog [Homo sapiens],
            full insert sequence.
ACCESSION   AK084780
VERSION     AK084780.1  GI:26351280
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
  PUBMED    10349636
REFERENCE   2  (bases 1 to 3617)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
  PUBMED    11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Ozawa,Y., Izawa,T., Tanaka,T., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Kira,A., Watahiki,M.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
  PUBMED    11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Saito,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
  PUBMED    11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3617)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Nunazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp;
            URL:http://genome.gsc.riken.go.jp/; Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3617
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:D330041I19"
                     /clone="D330041I19"
                     /db_xref="taxon:10090"
                     /tissue_type="heart"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="13 days embryo"
     CDS             <1..>3617
                     /note="unnamed protein product; TITIN homolog [Homo
                     sapiens] (SPTR|CAD12456, evidence: FASTY, 97.2%ID,
                     putative)"
                     /codon_start=3
                     /protein_id="BAC39277.1"
                     /db_xref="GI:26351281"
                     /translation="NARGTKGPKSRPSKPTVAMDFIAPPGKPQNPRVTDTTRTSVSLAW
                     SVPEDEGGSKVTGYLIEMQKVDQREWTKCNTTPKIREYTLTHLPQGAEYRFRVLACN
                     AGGPGEPAEVPGTVKVTEMLEYPDYELDERYQEGVFVRQGGVIRLTIPIKGKPPPVCK
```

```
Alignment Scores:

Pred. No.:                  0.172
Score:                    121.00
Percent Similarity:        35.00%
Best Local Similarity:     20.71%
Query Match:                3.07%
DB:                        11
```

```
BASE COUNT    1101 a    860 c    892 g    764 t
ORIGIN
```

ORIGIN sequence:

```
WTKEGQDISKRAMIATSETHTELVIKEADRNDSGTYDLVLENKCGKKTVYIKVVIGS
PNPEGPLEYDDIQARSVRVSWRPPADDGGADILGYIILERREVPKAAWTIDSRVGT
SLVVKLKENVEVHFRVSAENQFGISKPLKSEEPVIPKTPLNPPEPPSNPPEVLDVTK
FRIIAQNDVGLSETSPASEPVVCKDPFDKPSQPGELEILSISKISVTLQWEKPECDGG
KEILGIYWYERQSGDSANKKSNKERIKDRQFTIGGLLEATEYEFRVAENETGLSRPR
RTAMSVKTRLTSGEAPGVRKEMADVTVKLGEANAQLSCQIVGRPLPDIKWYRFGKELIQ
SRKYKMSSDGRTHHTLTVMIDEQDEGVYTCVATNEVGEVESSSKLLQAAPQFHPGYP
LKEKYYGAVGSTLRLHVHYIGRPVPAMTWFHGQKKLLQNSEKITIENTEHYTHLVMKNV
QRKTHAGKYKVQLSNAFGTVDAATLDVEIQDKPDKPTGPIVIEALLKNSVVSWKPPAD
DGGSWITNVVEKCEAKEGAEWQLVSSALSVTTCRIVNLTENAGYFRVSAQNTFGIS
EPLEVASIVILKSPFEKPGVGKPTITAVTKDSCCVVAWKPPASDGGAKIRNYLERRE
KKQNKWIAVTTEETRETVFSVQNLIEGLEYEFRVKCENLGGESEWSEISEPVTPKSDV
PIQAPHFKEELRNLNVRIQOSNATLVCKVTYGHPKPIVKWYRQGKEIIADSLKYRIQEFK
GGYHQLIIASVTDDDATVYQVRAIRNQGGSVSGTPASLEVEVPAKIHLPKTLEGMGAVHA
LRGEVVSIKIPFSGKPDPVITWQKGQDLIDNGHRYQVIVTRSFTSLVFSNGVERKDAG
FYVVCAKNRFGIDQKTVELDVADVPDPPRGVKVSDVSRDSVNLTWTEPASDGGSKVTN
YIVEKRETAERWLRVGQARETRYTVINLFGKTSYQFRVIAENKFGLSKPSEPSEPTV
T"
```

```
US-09-494-297-2 (1-757) x AK084780 (1-3617)

                        Length:        3617
                        Matches:        116
                        Conservative:    80
                        Mismatches:     196
                        Indels:         168
                        Gaps:            29
```

```
Qy    272  ProThrLysProProThrProGlyAspProProMetProProAsnGlnProGln------  289
           ||:|:||||||:|||||||                                    ::||||||
Db     33  CCTTCCAAACCCACCGTTGCCATGGATCCAAATTGCTCCACCAGGAAAACCACAAAATCCA  92

Qy    289  ------                                                        289
Db     93  AGAGTTACAGATACAACAAGGACTTCAGTCTCCCGGCCTGGAGTGTTCCGGAAGATGAA   152

Qy    289  ------                                                        289
Db    153  GGTGGATCTAAAGTCACAGGGTACTTGATCGAAAATGCAAAAATGATCAACGTGAATGG   212

Qy    290  ------ThrThrSerValLeuIleArgLysTyrArgIleGlyAspTyrSerLys        305
           ||:|||                      ||::|||:||::|||  ::|
Db    213  ACCAAATGTAACACACCCCAACCAAGATTCGGGACTATACCCTA------ACACAC       263

Qy    306  LeuLeuGlnLysAlaThrLeuGlnLeuThrGlyLysAspAsnValAsnSerPheGlnAlaArg  325
           |||  ::||||||||||                |||||||||
Db    264  CTACCTCAGGGCGCT------------------GAATACAGATTCCGT               293

Qy    326  ValPheSerSerAsn------AspIleGlyArgIleGluLeuSerAspGlyThrThr     343
           |||        ::||:       ||||||||       |||||||||
Db    294  GTCCTAGCCGTAAAGCCTGGGACCTGGAGAACCTGCTGAGGTA----CCAGGAACCGTC  350

Qy    344  ThrLeuThrGluLeuAsnSerProAlaGlyThrTyrSerIleAlaGluProIle------  360
           ::|||||||||||||||                |||||||          ||||||
Db    351  AAAGTCACAGGATGCTGGGAATACCCTGACTATGAACTTGATGAGGATACCAGGAAGGT   410

Qy    361  ThrPheLysValGlyLeuAlaGlyLysValGlyTyrThrIleIleAspGlyLysGlnIleGluAsn  380
           |||            ||:           ||:  :|||    |||      |||
Db    411  GTCTTTGTTCGACAAGGTGGGTCATCAGGCTTACCCATCCAAATCAAAAGGAAAACCATTC  470

Qy    381  ProAsnLysGluIleValGluProTyrSerValGluAlaTyrArgAsnAspPheGluGluPhe  400
           |||            :|:|||::|::|                ||:
Db    471  CCA------------GTATGTAAAGTGGCAAAGGGAGAAGGCCAAGACATTAGTAAGGCTGCC  518

Qy    401  SerValLeuThrThrGlnAsnGlyValThrAlaLeuLysPheTyrTyrAla----------  417
           :|:       |||||||||||            ||:           ::|
Db    519  ATGATTGCAACTACGACGACGACGACTGAGCTTGTCATCAAAGAGCCAGACAGAAACGAT  578

Qy    418  AsnGlySerSerGlnValValTyrCysPheAsnAlaAspLeuLysSerProProAspSer  437
           ||:||:|||::|||           |||           ::|           :||
Db    579  TCTGGAACTTACCAACTTGGTTCTG----------------------------GAAAAC  608

Qy    438  GluAspGlyGlyLysThrMet-----------------------------ThrProAspPheThr  449
           :::      ||||||                                    ||:
Db    609  AAATGTGGCAACAGAAACCGTTTACATTAAGATTAGGAGTCCCAACACTCCA             668

Qy    450  ThrGlyGluValLysTyrThrHisIleAlaGlyTyrArgAspLeuPheLysLysTyrArgValLys  469
           :::|||||||||||||||                |||||||          :::
Db    669  GAAGGCGCCACTGGAATACGGATGACATACAAAGCCCGTGTCTGAAGGTGCAGCTGGAGACCT,728

Qy    488  LysGly--------------------------------------TyrArgGlyLysGlyGlnAlaIleIleGluTyrSerGly  501
           |||                                          :::|||           |||
Db    729  CCTGCTGACGATGGTGGTGCTGACATCCTGGCGTTCAGCAATGGC----LysLysValIleGlu  788

Qy    502  LeuThrGluThrGlnLeuLeuArgAlaAlaAlaThrGlnLeuAlaIleTyrArgTyrPheThrAspSer  521
           |||||||                    ||||:||||  |||:||:           |||
Db    789  AAGGCTGCCCTGGTATACCATTGACCTCCAGAGTCCGAGGCCACATCTCTGGGTGGTGAAAGGC  848

Qy    522  AlaGluLeuAspLysAspLysLeuLysLysAspProTyrHisGlyPheGly-----------AspMet  538
           |||||||:|||||                        ||::|||                |||
Db    849  CTCAAAGAAAC----------------------GTGGAATACCATTTCCGTGTCTCA       884

Qy    558  ProProGlnLeuThrAspPhePheIleProAsnAsnAsnLysTyrGlnSerLeu  577
           ||||||:|||||                |||||||:|||
Db    978  CCTCCTCAAAGA------GTTACAGAGACAACCAGAGTCCGAGGCCAC         1010

Qy    578  IleGlyThrValTrp--------------HisProGluAsp------------LeuValAspIleLeuArg  592
           |||                                          ||:
Db   1011  GTTAGCTTGTCATGGTCTCGACCTAAAGATGACGGGGCGGGCTCTCGAGTGACGGGCTACTAC  1070

Qy    593  MetGluAspLysLysGluValIleProSer-------ValThrHisAsn-----------Leu  606
           :::|||                |||      |||:       |||||||
Db   1071  ATTGAGCGCCAAGAGACAATCCCACTGACAAGTGGGTCAGACACAACAAAACCCAGATAACC  1130

Qy    607  ThrLeuArgLysThrValThrGlyLeuAlaGlyAspArgThrLysAspPheHisPheGlu  626
           ||||||||||||||||||||                ||:          |||
Db   1131  ACCACGARGTACACTGTCAACGGCCGCCCTTCCGAT------GCTGAGTACCAGTTCCGC  1184

Qy    627  IleGluLeuLysAsnAsnLysGlnGluLeuLeuSerGlnThr------------  640
           :::|||||||                ||:           |||
Db   1185  ATCATCGCACAGAATGACGTTGGC----CTTAGTGAGACCCCTGCTTCTGAACCT  1238

Qy    641  ---ValLysThrAspPheAsnLeuGluPheLysAspGlyLysAlaThrIleAlaAsnLeu  659
           |||          |||:::                ||:          |||
Db   1239  GTTGGTTTGCAAAAGTCCCTTGTTGATAAGGCCAACAGCCCAACCAGCGAGATTCTCTCG  1298

Qy    660  LysHisGlyGluSerLeuThrLeuGln---GlyLeuProGlu--------------  672
           ||:       :||:::||||||||||    ||:
Db   1299  ATATCCAAAGACAGTGCCACCCTCCAGTGGGAGAAACCCGAATGTGATGGGTGGCAAAGAA  1358

Qy    673  -----GlyTyrSerTyrLeuValLysGluThrSerGluGlyTyrTyrLysLysVal  690
           ||||||                |||:::          :||:::          |||
Db   1359  ATCCTTGGATACTGGGTTGAATACAGCAGTGCTGGAGACAGTGCTTGGAAG-----AAGAGC  1415

Qy    691  AsnSerGlnGlyValAlaAsnAlaThrValSerLysThrGlyIle------------ThrSer  707
           |||          :::                ||:          |||
Db   1416  AATAAGGAACGGGATCAACAGAGACAGTTCACCATCGGCGGTTTGCTAGAAGCTACTGAA  1475

Qy    708  AspGluThrPheLeuAlaPheGluAsnAsnLysGlnProValValProThrGlyValAspAsp  727
           |||          ||:|::                ||:|||
Db   1476  TATGAGTTCCGAGTATTTGCTGAGAACGAG-----ACTGGGCGTCAGCCGA           1520
```

```
LOCUS       AK031936                2138 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male medulla oblongata cDNA, RIKEN full-length
            enriched library, clone:6330501K11 product:phosphatidylinositol 3
            kinase, regulatory subunit, polypeptide 3 (p55), full insert
            sequence.
ACCESSION   AK031936
VERSION     AK031936.1  GI:26082605
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
  PUBMED    10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
  PUBMED    11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
  PUBMED    11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
  PUBMED    11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  REFERENCE   6  (bases 1 to 2138)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
```

```
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2138
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6330501K11"
                     /sex="male"
                     /tissue_type="medulla oblongata"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    313..1697
                     /note="phosphatidylinositol 3 kinase, regulatory subunit,
                     polypeptide 3 (p55) (MGD|MGI:109277, GB|S79169, evidence:
                     BLASTN, 99%, match=2138)"
     polyA_signal    2113..2118
                     /note="putative"
     polyA_site      2138
                     /note="putative"
BASE COUNT      612 a    427 c    543 g    556 t
ORIGIN
```

```
Alignment Scores:

Pred. No.:          0.0967
Score:            120.00               Length:              2138
Percent Similarity:  36.00%           Matches:               93
Best Local Similarity: 19.58%         Conservative:          78
Query Match:          3.04%           Mismatches:           161
                                      Indels:               143
                                      Gaps:                  24

DB:

US-09-494-297-2 (1-757) x AK031936 (1-2138)

Qy     271 ValProThrLysProGlnThrProGlyAsp-ProProMetProProAsnGlnProGlnTh 290
               |||| |||||||||   |||||||||     ||| ||  | :::
Dd     419 CTTCCACCAAAGCCACCTAAGCCCAAATGACTCCACGATGCACAAATGGAATGAAGGACAG 478

Qy     290 r---ThrSerValLeuIleArgLysTyrAlaIleGlyAspTyrSerLysLysLeuLeuGluGl 309
               ||| :::            ||||:::            |||:::
Dd     479 TTTCATTTCTCTTCAAGATGCGAGAGTGGTTACTGGGGAGATATTTCCAGG----------- 527

Qy     309 yAlaThrLeuGlnLeuThrGlyAspAsnValAsnSerPheGlnAlaArgValPheSerSe 329
               ||| :::            ||| ::            ||| :::
Dd     528 -------GAAGAGGTAAATGACAAATTTGCGGGACATGCCCAGATGGTACCTTCTTAGT 577

Qy     329 rAsnAspIleGlyGluArgIleGluLeuSerAspGlyThrTyrThrLeuThr--------- 346
               |||:::::|            |||  ||||||||||||
Dd     329 ...
```

```
Db   578  TCGTGATGCCTCAACGAAAAATGCAG-------                  GGGGATTATACATTGACTTTGAGGAA  628
Qy   347  --------------GluLeuAsnSerProAlaGlyTyrSerIleAlaGl  358
Db   629  GGGAGGAAAATAAATTAATAAAGATCATATCATCGGGAAGGTAAATATGGCTTCTCTGA  688
Qy   358  uProIleThrPhe------------  362
Db   689  GCCCTGACGTTTACTTCTGTGTGGGGAGCTTATTAACCACTACCACCACGAGTCTCTCGC  748
Qy   363  --------LysValGluAlaGlyLysValThrThrIleIleAspGlyLysGln--  377
Db   749  TCAGTACAATCCCAAACCTGACGTGAAGCTGACGTACCCAGTATCCCAGTATCCAACAGGA  808
Qy   378  ------IleGluAsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAs  395
Db   809  TCAGTTGGTAAAAAGAAGATAACATTGATGCAGTAAAAAATCTGCAGAGAGTTCCACTC  868
Qy   395  nAspPheGluGluPheSerValLeuThrThrGlnAsnTyrYrTyrTyrAlaLy  415
Db   869  TCAGTATCAGGAGAAGAGC-----------  AAAGAGTATGACAGGCTGTAT-----GA  910
Qy   415  sAsnLysAsnGlySerSerGlnValValTyrCysPheAsnAlaAspLeuLysLysSerProPr  435
Db   911  AGAGTACACAAGGACGACATCACAGAAATA  938
Qy   435  oAspSerGluAspGlyGlyLysThrMetThrProAspPheThrThrGlyGluValLysTy  455
Db   939  -------  CAAATGAAGAG  949
Qy   455  rThrHisIleAlaGlyArgAspLeuPheLysTyrTyrValLysProArgAspPr  475
Db   950  GACTGCCATT-------  GAAGCCTTAATGAAACAATTAAATATTGAGGAGCAGTG  1000
Qy   475  oAspThrPheLeuLysHisIleLeuLysSerValIleGluIleLysGlyTyrArgGlyLysGlyGl  495
Db  1001  TCATACCCAAGAACAACACAGTAAAGACTATATCGAGCGC---TTCGAGAGAGGGGAA  1057
Qy   495  nAlaIleGluTyrSerGlyLeuThrGlnLeuArgGlnAlaAlaThrGlnLeuAlaIl  515
Db  1058  TGAGAAGGAGATCGAGCGAATTATG-----------  1082
Qy   515  eTyrTyrProThrThrAspSerArgAlaGluLeuAspLysSerAspLysSerAspTyrHisGlyPh  535
Db  1083  ATGAATTATGATAAATTGAAATCA------  CGTCT  1111
Qy   535  eGlyAspMetAsnAspSerThrLeuAlaValAlaLysIleLeuValGluTyrArgAlaGlnAs  555
Db  1112  TGGTGAGATTCATGATAGCACAAACTCCGTCTTGAGCAGTGAAGAACAAAGCTTTGGA  1171
Qy   555  pSerAsnProProGlnLeuThrAspLeuAspPhePheIleProAsnAsnAsnLysTyrGl  575
Db  1172  CAACCGG-------  GAAATAGAT------  1198
Qy   575  nSerLeuIleGlyThrGlnTrpHisProGluAspLeuAspLeuGluLeuLysSerArgMetGluAs  595
Db  1199  TAGCATC-------  AAACCCGACTTGATCCAGCTGCGTAAGACCCGGGA  1240
Qy   595  pLysLysGluGlyValIleProValThrThrAspLysThrValThrGlyLye  615
Db  1241  T---CAGCACCTTGTATGGCTCAATCACAGA  1268
Qy   615  uAlaGlyLysAspArgThrLysAspPheHisGluGluLeuLysSerAsnAsnLysGlnGl  635
Db  1269  ----GGAGTGAGCAGAGGCGCCTCGAATGCCTGGGGATCAAGAATGAGGAGACTCAGA  1324
Qy   635  uLeuLeuSerGlnThrArgValLysThrAspLysThrAsnLeuGluPheLysAspGlyLysAl  655
Db  1325  T---GAAAGCTATTTATTCAAATGAGGAAGATGAAGAACCTGCCCGCCAATATGATGACAAAAAC  1381
Qy   655  a--------        ThrIleAsnLeuLysHisGlySerLeuThrLeuGlnGlyLeuPr  671
Db  1382  CTGGGTTTGTTGTGGGAGGATATCAACCGGAGTACAAGCAGACAGGAGGACTTG---CTTTATGGGAAACC  1438
```

```
Qy   671  oGluGlyTyrSerTyrLeuValLysGluThrAspSerGluGly--TyrLysValLysVa  690
Db  1439  AGATGGT--GCAATTCTTAATTCGTGAGAGTAGCAAGAAAAGGATGATTACGCTTGTTCTGT  1495
Qy   690  lAsnSerGln---GluValAlaAsnAlaThrValSerLysThr  703
Db  1496  GGTTGCAGACGCGGGAAGTGAACGCACTGTGTCATCTACAGCACG  1538
```

RESULT 5

AK053071

LOCUS   AK053071   3959 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION   Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched
             library, clone:D930043C02 product:MEMBRANE-ASSOCIATED NUCLEIC ACID
             BINDING PROTEIN (FRAGMENT) homolog [Homo sapiens], full insert
             sequence.
ACCESSION   AK053071
VERSION   AK053071.1 GI:26095552
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
             Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
             Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
             Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
             Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
             Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
             Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
             Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
             Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
             Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
             Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
             Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
             Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
             Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
             Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
             Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
             Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C.A., Wilming,L.,
             Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
             and Hayashizaki,Y.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660

```
PUBMED     5
REFERENCE  The FANTOM Consortium and the RIKEN Genome Exploration Research
AUTHORS    Group Phase I & II Team.
TITLE      Analysis of the mouse transcriptome based on functional annotation
           of 60,770 full-length cDNAs
JOURNAL    Nature 420, 563-573 (2002)
PUBMED     12466851
REFERENCE  6  (bases 1 to 3959)
AUTHORS    Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
           Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
           Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
           Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
           Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
           Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
           Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
           Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
           Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
           Sogabe,Y., Tagami,M., Tagami,A., Takahashi,F., Takaku-Akahira,S.,
           Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
           Muramatsu,M. and Hayashizaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
           URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
           Fax:81-45-503-9216)
COMMENT    cDNA library was prepared and sequenced in Mouse Genome
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in RIKEN.
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues.
           Please visit our web site for further details.
           URL:http://genome.gsc.riken.go.jp/
           URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3959
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:D930043C02"
                     /db_xref="taxon:10090"
                     /clone="D930043C02"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="15 days embryo"
     misc_feature    241..3784
                     /note="MEMBRANE-ASSOCIATED NUCLEIC ACID BINDING PROTEIN
                     (FRAGMENT) homolog [Homo sapiens] (SPTR|Q9HBD2, evidence:
                     FASTY, 96.1%ID, 99.9%length, match=3128)
                     putative"
BASE COUNT     1168 a    923 c    866 g   1002 t
ORIGIN
```

```
US-09-494-297-2 (1-757) x AK053071 (1-3959)

Alignment Scores:
Pred. No.:                 0.256
Score:                   120.00
Percent Similarity:       31.94%         Length:         3959
Best Local Similarity:    18.80%         Matches:         166
Query Match:               3.04%         Conservative:    116
DB:                        11             Mismatches:      321
                                          Indels:          281
                                          Gaps:             40
```

```
Qy      1 MetLysLysThrArgPheProAsnLys-----LeuAsnThrLeuAsnThrGlnArgVal  18
          |||||||||||                   |||||      |||    |||
Db   1557 CTTGAAAAGTACCGATTAAGGAACAAAAAGATGAGCGCGACTGTAAGACAACATTTCTCTT 1616

Qy     19 LeuSerLysAsnSerLysArgPheThrValThrLeuValGlyValPheLeuMetIlePhe  38
          |||:::||||                    |||||||||||||
Db   1617 CTGAATAAAGTTGGTGTAAACACCACTGTCACCAACCACCAGCC----------------- 1658

Qy     39 AlaLeuValThrSerMetValGlyAlaLysThrValPheGlyLeuValGluSerThr  58
          |||      |||:::|||||||:::|||:::|||      ||| :::|||:||||||||
Db   1659 GGAAATGTCATTTCTGTCATAGGAAGTACTGAAACAACTGGGAAAATTGTTGAAGTACA 1718

Qy     59 ProAsnAlaIleAsnProAspSerSerSerSerGluTyrArgTrpTyrGlyTyrGluSerTyr  78
          |||||| |||    :::||||||
Db   1719 AATGGAATTTCAAATTACAGAAAGCAGT---------------------------------- 1745

Qy     79 ValArgGlyHisProTyrTyrLysGlnPheArgValAlaHisAspLeuArgValAsnLeu  98
                              ||||||
Db   1745 ------------------------------------------------------------ 1745

Qy     99 GluGlySerArgSerTyrGlnValTyrCysPheAsnLeuLysLysAlaProLeuGly  118
          ||| |||      ||| :::      |||
Db   1746 ------------GTTTCTCAGCTAATCCCACGAGGC------------------------- 1769

Qy    119 SerAspSerSerValLys-----LysTrpTyrLysHisAspGlyIleSerThrLys  136
          :::||||||||||:::|||:::                 |||||||:||
Db   1770 ACTGACAGCTCCAGTAAGGACTTGGAGACTGTGAGAAAGTGGTGGCACTAAT 1829

Qy    137 PheGluAspTyrAlaMetSerPro---------------------ArgIleThrGlyAspGlu  150
                                                        :::|||||||
Db   1830 GCTCAGAATGCTGGCCCTTGCAGAGTCCGTGTCTGAAATAAAATTGTTCTCCACC 1889

Qy    151 LeuAsnGlnLysLeuArgLaValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyIle  170
          ||| |||||||||||||||       |||       ||||||
Db   1890 AAGACTCCTGTAAGTAATGGCAGCAGCTACCTCCGGTGGGCCCCTTCAAATTTGGAACAGAG 1949

Qy    171 MetGluGlyLeuGluProLeuAsnAlaIleLeuArgValThrGln----------  184
          :::                       |||
Db   1950 CTGAATTCCCTGCCTCCCAAAATCCAGCCCATTTCTAACTAGAGTTCCAGTATATCCTCAG 2009

Qy    185 -----GluAlaValTrpPyrTyrSerAspAsp----------AsnAlaProIleSerAsnPro  199
                         |||             |||
Db   2010 CATTCTGAAAGCATTCAGTATTTTCAAGAACCCAAGGACTCAGATACCTTTTTGAAGTCCCA 2069

Qy    200 Asp-----  200
          ||
Db   2070 CAATACCCACAAACAGGATACTACCCACCACCTCCAACGGTACCAGCTGGTGTGACTCCC 2129

Qy    201 -----GluSerPheLysArgGluSer----------GluSerAsnLeuValSerThrSer  215
                         |||      |||       |||||||
Db   2130 TGTGTTCCTCGCTTTGTTGAGGTCCAGTAAGTTCCAGAGTGTCCTCCCTC---------- 2177

Qy    216 GlnLeuSerLeuMetArgGlnAlaLeuLysGlnLeuIleLeuIleAspProAsnLeuAlaThrLys  235
          |||                                               |||
Db   2177 ------------------------------------------------------------ 2177

Qy    236 MetProLysGlnValProAspPheProPheGlnLeuSerIlePheGluSerGluAspLysGly  255
              ||| ::: |||                   ||| ||| |||:::
Db   2178 CCACCTGCTTCCATGCCATATGCTGATCATCAATTCCCCCTCCGGAATCGAATG 2237

Qy    256 AspLysTyrAsnLysGlyTyrGlnAsnLeuLeuSerGlyLeuValProThrLysPro  275
          :::                                  |||             |||||
Db   2238 AATTCT-------------------------------TCTTCCTTACCAACCT 2258

Qy    276 ProThrProGlyAsp----------ProProMetProProAsnGln----ProGln  289
          |||                     ||||||||||||||||||||
Db   2259 CCTCCTCCCGCACGATATGGACCAGTTCCCACCCCAGTAACCCTGGAATG 2318

Qy    290 ThrThrSerValLeuIleArgLysTyrAlaIleGlyAspTyrSerLysSerLeuLeuGlyTyr  309
          |||                        :::||| |||
Db   2319 TATGACAGCAGGCGGCCGCATCGGCGCCCAGCTATGTACCAACGAGATGACATTATTAGAAGC 2378

Qy    310 AlaThrLeuGlnLeuThrGlyAspAsnValAsnSerPheGlnAlaAlaArgValPheSerSer  329
          |||::::|| |||             :::
Db   2379 AATTCTTTA----------CCTCCAATGGATGTTGGATGTTGATGCACTCAATCTGTCTCACAGACA 2426

Qy    330 AsnAspIleGlyGluArgIleGluLeuSerAspGlyThrTyr-----  343
          ||||||           |||
Db   2427 TCT---------TTACGGGAAAGAATACAACTCATTAGATGCGATTATTATTCAGTGCCA 2483
```

```
Qy    344  ------------------------ThrLeuThrGluLeuAsnSerProAlaGlyTyr-------------------  354
                                   |||  :::                           |||:::
Db   2484  CCGAATGACCCAAGGACAACTGTGCCTTGCCTTACCAAGGGAACCTTGTGGTCATTGAATTGAAGACC         2543

Qy    355  SerIleAlaGlyProIleThrPheLysValGluAlaGlyValTyr-----------------------------  370
                 |||         ::::                           |||
Db   2544  AGTTGTGAGGAACAGTTAAGAAGAAAGCACAGATACAGTGGACACAGTATCATACCCAGAAA              2603

Qy    371  ------------ThrIleIleAspGlyLysGlnIleLeuAsnProAsnLysGluIle-----  385
                       |||:::        |||                  |||
Db   2604  ACACCTGTCTCTTCAACCTCTTCCTGTGGCACCACAGTCACCACCACCACCTTCCTCTA        2663

Qy    386  ValGluProTyrSerValGluAlaTyrArgAsnAspPheGluPheSerValLeuThrThr-----  405
                :::|||::::|         ::::::|||||||      :::
Db   2664  AATTCAAATGCTGTATTGATTGATGGACCTGGACAGTGGGGATGTGGAGTAAAGAGAAGAGTGGCAAT----  2870
```

```
Qy    406  GlnAsnTyrAlaLysPhe--------------------------TyrTyrAlaLysAsnLysGly-----  419
                |||||||||                          |||
Db   2700  GTGAGTGGCGCCAAAATTGAGGAAGACCAATCTTTCCCATTATTCTCCTGGTCTTGTGGT             2759

Qy    420  SerSerGlnValValTyrCysPheAsnAlaAspLeuLysSerProProAspSerGluAsp-----  439
                |||       :::            |||                  |||
Db   2760  ACC------ATAGGCTCTTGTATAAAATGCCATTGAATTCGGACCCCAAAGACGTAATTGCT        2813

Qy    440  GlyGlyLysThrMetThrProAspPheThrThrGlyGluValLys---TyrThrHisIle  458
                |||            ::::|||                       |||
Db   2814  CGATCCTCCGAACAGGT----------------------------TACCATACA          2870

Qy    459  AlaGlyArgAspLeuPheLysTyrArgHisValLysProArgAspThrArgAspThrPhe  478
                |||||||||                    :::  :::::        |||
Db   2871  ------------TTATTGAAGCTCAGAGAGTGCAACTAGCGTCCCATACGGCTATCA----  2909

Qy    479  LeuLysHisHisLeuLysValIleGluLysGlyTyrArgGluLysGlyGlnAlaIleIleGlu  498
                                    |||                 :::
Db   2910  ------------ATTATTCCCTTTAGTGACGGACCCATTATCTCAAAAATGGGGTGCAATTTCC  2960

Qy    499  TyrSerGlyLeuThrGluThrGlnLeuArgAlaLeuaThrGlnLeuIleAlaIleTyrTyrPhe  518
                |||                               |||
Db   2961  CGATCCTCCGAACAGGT-----------------------TACCATACA             2987

Qy    519  ThrArgSerAlaGluIleLeuAspLysAspLys--------------------------LeuLys  530
                |||                                |||
Db   2988  ACAGATCCGGTCCAGGACAACTGCTTCCCAGAGAAGTGCAACTAAGCCCATCAGCGTATCA      3047

Qy    531  AspThrTyrHisIleGlyPheGlyAspMetAsnAspSerThrLeuAlaValAlaLysIleLeuVal  550
                |||                                         :::
Db   3048  GATTATTGTCCCTTAGTGAAGCTGTCGATTCAAGGTGGACT-------------------------  3089

Qy    551  GluTyrArgAlaGlnAsp---SerAsnProProGlnLeuThrAspLeuAspPhePheIlePro  569
                |||                |||                     ||||||
Db   3090  TCAATATGGCGAACGAGATGCTACATCAGCACACTATATTGAAACGGGACAGAATTCATTGTT      3149

Qy    570  AsnAsnAsnLysThryrGlnSerLeuIleGlyTyrThrGlnTrpHisProGlu-------AspLeu  587
                :::                                       |||         |||||||
Db   3150  ACTGAT-----------------TTATCTGGTCATAGAAGGCATTCTAGTACCGGAGACCTT      3194

Qy    588  ValAspIleIleArgMetGluAspLysLysGluValIleProValThrHisIleAsnLeuThr  607
                ||||||                           ::::::::::|||
Db   3195  TTGACCATTGAACTTCAGCAGGCCAAG------------------AGTAACTCATTACTT      3236

Qy    608  LeuArgLysThrValThrArgGlyLeuAla-------------GlyAspArgThrLysAsp  622
                ||||||:::::                                  :::
Db   3237  CTTCAGAGGGAAGCTAATGCTCTGGCCATGCAACAAAAAGTGGAATTCCCTGGATGAAGGC  3296

Qy    623  PheHisPhe----------------------GluIleGluLeuLysSerAsnLysSerGlnGlu  635
                |||                              |||||||||||||||::||:|||
Db   3297  CGTCACCTTACTTTAAATCTTCTAAGCAAGGAAAATTGAACTAAGAAATGAGAGAAGAATGAC      3356

Qy    636  LeuLeuSerGlnThrValLysThrAspLysThrAsnLeuGluPheLysLysAspGlyLysAla  655
                |||:::        |||               :::         ::|||
Db   3357  TACACAGAAGACACAGTGTAGATACA---AAGCCTGATAGGGATAATTGAG-----------  3401

Qy    656  ThrIleAsnLeuLysHisGlyGluSerLeuThrLeuGlnGlyLeu---------Pro   671
                |||                                            |||
Db   3734  AAAAACGAC 3742
```

```
Db   3402  -----------------------------------------TTAGAGCTTTCAGCCCTTGAITACTGATGAACCT  3434
                                                       |||     |||            |||
Qy    672  GluGlyTyrSerTyrLeuValLysGluThrAspSerGluGlyTyrLysGlyLysValAsn  691
                :::|||   |||                   :::::|||
Db   3435  GATCGGCCAGAGTGAACAAATTGAAGAAATTTGGACATACAACAACTGGTATCAGTTCTCAA  3494

Qy    692  SerGlnGluValAlaAsnAlaThrValSerLeuTyrGlyIleThrSerAsp--------  708
                ::::::|||   |||          ::  |||::::
Db   3495  AATGACCAGTTCCTCAATTGGACACAGCAGTGGAAAATGGTCA-TCCAGCCCACCAGCACCA  3553

Qy    709  -----------GluThrLeuAlaPheGluAsnAsnLysGlyLeuProValValPro  722
                           |||          :::::::|||   :::
Db   3554  GAAGGACCCAGGAGAAAAGCCGAAGAGAGAGTTTAGGGAGGGAGGAAAAAACCAATTCTCCG  3613

Qy    723  ThrGlyVal-------------------AspGlnLysIleAsnGlyTyrLeuAlaLeuIle  736
                |||                      |||:::                |||
Db   3614  GTAACATCTTGCTTAGCCGACGCCACGTCCAAATGTCTATTAGCAGTGCAAGCTGCCTCCC      3673

Qy    737  ValIleGluAlaGlyIleLeuSerLeuGly-----IleTrpGlyIleHisThrIleArgIleArg  754
                :::::||:|                     |||                    |||
Db   3674  ATCACCACATCTGTCAGTGTTGGCCAACCTCATTCTGAAAACTCACGTTATGTCTGAAGAT     3733
```

```
REFERENCE   5  (bases 1 to 4594)
  AUTHORS   Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
    TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
   MEDLINE  21085660
   PUBMED   11217851
REFERENCE   6  (bases 1 to 4594)
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
    TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   7  (bases 1 to 4594)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
    TITLE   Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
   COMMENT  cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..4594
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:9830138J03"
                     /db_xref="taxon:10090"
                     /clone="9830138J03"
                     /sex="male"
                     /tissue_type="bone"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             26..2347
                     /note="unnamed protein product: putative
                     transcription elongation factor B (SIII), polypeptide 3
                     (110KD) [MGD|MGI:1351315, GB|NM_013736, evidence: BLASTN,
                     99%, match=2537]"
                     /codon_start=1
                     /protein_id="BAC29497.1"
                     /db_xref="GI:26331534"
```

```
     polyA_signal    4574..4579
                     /note="putative"
     polyA_site      4594
                     /note="putative"
ORIGIN
BASE COUNT     1278 a   1091 c   1108 g   1117 t
```

```
                     /translation="MAAESALQVVEKLQARLAANPDKKLLKYLKKLSILPITVDILV
                     EFGVGKTVNSFRHEQVGNFARDLVAQWKKLPVVERNSEAEQDDFEKNNSRKRPRDAL
                     QREEELEGNYQESWKSGSSRSYSPEHRQKKHKKLSEPEPHKVAHSHEKRDERKRCHK
                     VSPPYSSDPDESSDYGHVQSPPSPHQMYTDLSRSPEEDQEPIISHQKPGKVHSNTFQ
                     DRLGVSHLGEQGKGAVSHHKQHRSSHKEKHPADAREDEKISAVSREKSHKASSKEESR
                     RLLSGDSAKEKLPFSSVVKKDKDREGSSLKKFKFSPALDVASDNHFKKPRHKDSEKAKSD
                     KNKQSVDGVDSGRGTGDPLPKAKEKVNHLKAQEGKVRTNADGSKSAGPLHPKAEETDV
                     DDEFERPIMSFESYLSYDQPRKKKWKVVKTSSTALGEKGLKKKDSKSTSKNLNSAQKL
                     PKVNENKSEKLQPAGAEPTRPRKVPTDVLPALPDIPLPAIHANYRPLSLELIPSFQP
                     KRKAFSSPQEEEEAGFGRRMNSKMQVYSGSCAYLPKMMTLHQQCIRVLKNIDSIF
                     EVGGVPYSVLEPVLERCTPDQLYRIECNHVLIEETDQLWKVHCHRDEKEEPEEYES
                     WREMYLRLQDAREQORLRLLTNNIRSAHHANKPKGRQAKMAFVNSVAKPRDVRROEKF
                     GTGGAAVPEKVRIKPAPYTTGSSHVPASNSSSNFHSSPEELAYDGPSTSSAHLAPVAS
                     SSVSYDPRKPAVKKIAPMMAKITKAFKNRFSRR"
```

```
US-09-494-297-2 (1-757) x AK036592 (1-4594)

Alignment Scores:
  Pred. No.:                       0.477
  Score:                           118.50
  Percent Similarity:              35.25%          Length:           4594
  Best Local Similarity:           22.17%          Matches:           100
  Query Match:                      3.00%          Conservative:       59
  DB:                              11               Mismatches:        148
                                                   Indels:            144
                                                   Gaps:               23
```

```
Qy    99 GluGlySerArgSerTyrGlnValTyrCysPheAsnLeuLysLysAlaPhe------- 115
         ||||||||||                       :::|||||||||   |||
Db  1013 AAAGCCAAAATCTGACAAAAAACAAGCAGAGTGTAGATGGCGTGGACTGACGACGAGGACA 1072

Qy   116 ---ProLeuGlySerAspSerSerValLysLysTrpTyrLysLysHis---AspIleIle 133
         :::                          ||||||           |||||
Db   923 GAAGGCAGC------------------------AGTCTCAAGAGAAGAAGTTTCACCTGCC 958

Qy   134 SerThrLysPheGlu--------------------AspTyrAlaMetSerProArgIleThr 147
         |||     :::                        |||   |||   |||
Db   959 TTAGATGTTGCTTCAGACAACCACTTTAAAAAG-----CCCAAACACAAGGACTTCCGAG 1012

Qy   148 GlyAspGlyLeuAsnGlnLysLeuArgAlaValMetTyrAsnGlyHisProGln------ 165
         ||||||||      |||||           ||||::::         |||
Db  1073 GGAGACCCATTGCCCAAAGCCAAGGAGAAAGTTCCCAACCACCTGAAGGCTCAGGAGGGG 1132

Qy   166 ---------AsnAlaAsnGlyIleMetGluGlyLeuGluProLeuAsnAla------- 179
         ||||||::::|||             |||||            :::
Db  1133 AAAGTAAGAACTAACGCACAGATGGAAAAGTCAGCAGCC----CCGCTCACCCTAAAGCG 1186

Qy   180 -------------IleArgValThrGlnGluAlaValTrpTyrTyr 190
                                   |||                |||
Db  1187 GAGGAGACAGATGTGGACGATGAGTTTGACGGCCCACATGTCCTTCGAGTCCTTC 1246

Qy   191 SerAspAsnAlaProIleSerAsnProAspGluSerPheLysArgGluSerGluSerAsn 210
         |||       :::                       ||:::           :::
Db  1247 AGCTATGACCAGCCC--------------------CGGAAGAAAAAGAAGAAG 1279

Qy   211 LeuValSerThrSerGlnLeuSerLeuMetArgGlnAlaLeuLysGlnLeuIleAspPro 230
         :::||| |||           |||   :::         ||||||:::
Db  1280 GTTGTGAAAAACGTCCAGTACAGACACTTGGAGAAAAAAGGACTTAAAAAAGGACTCTAA 1339

Qy   231 ------------AsnLeuAlaThrLysMetProLysGlnValProAspAspPhe 244
                              |||       |||   |||
Db  1340 AGCACTAGTAAAAACTTGACACTCAGCTCAGAAATTGCCCAAA------------ 1381

Qy   245 GlnLeuSerIlePheLeuSerGluAspLysGlyAspLysTyrAsnLysGlyTyrGlnAsn 264
                       ::::|||:::|||   :::|||   :::|||
Db  1382 ------------GTCAACGAAAAACAAGTCAGAGAAG 1405

Qy   265 LeuLeuSerGlyGlyLeuValProThrLysSerPro------ 276
         |||       |||                    |||
                                                   Pro 276
```

```
Qy   1406  TTGCAGCCACCTGAGCCGCGAACCCAGCCGCCTCGAAAAGTCCTCTACTGATGTGCTGCCA 1465
                ||||                   ||| |||||:||||
Db    277  ThrProGlyAspProMetPro--------------------------ProAsnGln 287

Qy   1466  GCATTGCCAGACATCCCCTTACCTGCCAACTACCGCGCCCGCCTTCCTTCCCTT 1525.
                |||:|||||                         ||| :||:||||
Db          Seq primer: ADF.

Qy   1406  ProGlnThrProSerValLeuIleArgLysTyrAlaIleGlyAspTyrSerLysLeuLeu 307
                |||                    ::|||   ||||
Db    288  ProGlnThrProSerValLeuIleArgLysTyrAlaIleGlyAspTyrSerLysLeuLeu 307

Qy   1526  GAGTTGAGTTCCCTCCTTCAGCCAACCAAAGCGAAAAGCT------------ 1576
                ||||:||||                    ::|||||||
Db    308  GluGlyAlaThrLeuGlnLeuThrGlyAspAsnAlaAsnSerPheGlnAlaAlaArgValPhe 327

Qy   1577  GAAGAAGAAGAAGAGTGGGTTCACAGGACGCGAAATGAAATTCT--AAGATGCCAGGTGTAT 1633
                :::|||||||||                    ::::|||||||
Db    328  SerSerAsnAspIleGlyGlu------------------------Arg 335

Qy   1634  TCTTGGTTCCAAGTGTGCCTATCCTCCCCAAAATGATGACATTGCACCAGCAGTGTATCCGG 1693
                ||||                   |||
Db    336  IleGluLeuSerAspGlyThrTyrThrLeuThrGluLeuAsnSerProAlaGlyTyrSer 355

Qy   1694  GTG---CTTAAGAATAATATTGACTCATCTTTGAAGTAGGAGGA--GTCCCCTTATTCT 1747
                |||                    :|:||||
Db    356  IleAlaGluProIleThrPheLysValGlyLysValIleTyrThrIleIleAspGly 375

Qy   1748  GTTCTTGAACCTGTCTTGGAGAGTGTCACACCTGATCAGCTCTAC-------------- 1792
                |||||:|||                    |||
Db    376  LysGlnIleLeuGluAsnProAsnLysGlyIleValGluPro--------TyrSerVal 391

Qy   1793  ---CGGATAGAGGAAATGCAATCATCGTATATAATTGAAGAAACAGATCAGTTGTGGAAAGTT 1849
                :::|||||                    :::|||||
Db    392  GluAlaIleTyrAsnAspPheGluGluPheSerValLeuThrPheGlnAsnTyrAlaLysPhe 411

Qy   1850  CACTGTCACCGAGACTTTAAGGAGAGAAAAGCCAGAAGAGTATGAGTGTCTGGAGGGAGATG 1909
                :::||||||                    ::|
Db    412  Tyr---- 412

Qy   1910  TACCTTGAGGCTTCAGGACGGCCAGAGAGCACGGGTGTCGCGCTGCTCACAAACAACATCCCGC 1969
                |||                    :|:|||
Db    413  -----TyrAlaLysAsnLysAsnGlySerSerGlnValValTyrCysPheAsnAlaAsp 430

Qy   1970  TCTGCTCATCCCAATAAAGCCACAAAGGGACGACAAGCAAAGATAGGGCTTTTGTCAACTCGTG 2029
                ||||||||:||                    :::
Db    431  LeuLysSerProProAsp--SerGluAspGlyGlyLysThrMetThrProAspPheThrThr 450

Qy   2030  GCCAAGCCACCAAGAGAGTGTCCGAAGAAGGCCAGGAGAAGTTTGGAACCG------ 2078
                |||  ||| ||||||||:||||                    |||
Db    451  TyrSerVal 391

Qy   2079  -GGGGAGCAGCTGTTCCTGAGAAAGTCAGGA 2108
                ::|||||
Db          rGlyGlyValLysTyrThrHisIleAlaGly 460
```

```
BASE COUNT    311 a    50 c    136 g    207 t
ORIGIN
```

```
US-09-494-297-2 (1-757) x BM169720 (1-704)

Alignment Scores:
Pred. No.:
Score:              116.00        Matches:        704
Percent Similarity: 38.70%        Conservative:    47
Best Local Similarity: 20.69%     Mismatches:      86
Query Match:         2.94%        Indels:          74
                                  Gaps:            11
Pred. No.:           0.0464       Length:         704
                                  Mismatches:      86
DB:                  12           Gaps:            11
```

```
Qy    368  AGTTATAATTTTGATAGAAATGAAGATGATAGAGATGAGAAGTATCTTTC------------- 155
                |||:::                   :::
Db    111  LysValTyrThrIleIleAspGlyLysGlnIleLeuGluAsnProAsnLysGlyIleLeuGlu 387

Qy    408  TyrAlaLysPheTyrTyrAlaLysAsnLysAsnGlySerSerGlnValValTyrCysPhe 427
                |||                   :::|||||:||||:::
Db    156  --------------------CCTGATGGGGGAGGCTGTGCCGAAGTTGATAGATAACATA 194

Qy    388  ProTyrSerValGluAlaIleTyrAsnAspPheGluGluPheSerValLeuThrPheGlnAsn 407
                |||:::                   |||  |||||||
Db    111  AGTTATAATTTTGATAGAAATGAAGATGATAGAGATGAGAAGTATCTTTC------------- 155

Qy    368  LysValTyrThrIleIleAspGlyLysGlnIleLeuGluAsnProAsnLysGlyIleLeuGlu 387
                |||                   :::|||||:||||:::
Db     57  AAAAAAAAAAAAATTTAAGTGATAATGATAAAATCAGAAAAATGAAAGAAAAGAATTA----- 110

Qy    428  AsnAlaAspLeuLysSer---------------------------------------- 436
                |||||:::                   |||||:
Db    195  GAGGCAGACATTGGCTCTGATTTAGATAGTGTAGATGTCCCACGAAAGAAAAATGAC 254

Qy    437  SerGluAspGlyGlyLysThrMetThrProAspPheThrThrGlyGlnValValLysTyrThr 456
                |||:::||:||                   :::|||:::
Db    255  AGTGATAATGGTGAGAAAAATATTGATCCATCATTGATATCATCAGAAAATAGAATATCAAT 314

Qy    457  HisIleAlaGlyArgAspLeuPheLeuTyrThrValLysProArgAspThrAspProAsp 476
                |||                   -----|||
Db    315  GATATA------------------------AAAATTGATGATGAGACGATGAT 344

Qy    477  ThrPheLeuLysHisIleLeuLysLysValIleGluLysGlyTyrArgGlyGlyGlnAla 496
                :::                   ::|||||:::|||
Db    345  GATAATATAAATAACTTGTCAAAAATGACAAAAAAGAAAAATCTTTTGAAAATGAAGAGGTT 404

Qy    497  IleGluThrTyrSerGlyLeu-----ThrGluThrGlnLeuArgAlaAlaThrGlnLeuAla 514
                |||:::                   |:|||||||:|
Db          ProAsnGln
```

```
RESULT 7
BM169720
```

```
Qy    405  AAGGAGGCAGGCTTGGTATTGATGCTTCAAAAACTGAGTTGCAA----------------------- 446
           |||    :::||/|||      ::|||::|:|||||||||:|||::
Db    405  AAGGAGGCAGGCTTGGTATTGATGCTTCAAAAACTGAGTTGCAA----------------------- 446

Qy    515  IleTyrTyrPheThrAspSerAlaGlyLeuAspLysAspPtyrHisGly 534
           ||||||:|::||   ||||||:|:||||:|||||
Db    447  ------------AATAATATTGAATTAACAAAAGTATAAA--------------------------- 473

Qy    535  PheGlyAspMetAsnAspSerThrLeuValAlaValAlaLysIleLeuValGluTyrAlaGln 554
           ||||||||:     :::
Db    474  -----GATATGAATTTCGATTTGTTCGGAAAATACAAAAATG------------------------ 512

Qy    555  AspSerAsnProProGlnLeuThrAspLeuAspPhePheIleLeuProAsnAsnAsnLeuTyr 574
           ::|||||      |||    :::          |||  ||||||  |||||||||||||||||
Db    513  GAAGAAAATAAAGTGTAGATTCAGATGACGAGATAAATATTGAAGAAAAATAATTATTG 572

Qy    575  GlnSerLeuIleGlyThrGlnThrHis------------------ProGluAspLeu 587
           |||||||||||        |||||||                      |||||||:|||
Db    573  CAT--------GATTGGTATGTTAAAAAAAATGAGAATAATAAAAAGAAGAATGTT 617

Qy    588  ValAspIleIleArgMetGluAspPtyLysGluValIleProValThrHisAsnLeuThr 607
           ::|||:::||||      |||||||:||||::|||:||
Db    618  TTAAAGTGTTG---TTTAGTGATATGCAAAAATGTTTTACCTTTGGAAGAAAAATTTGACT 674

Qy    608  Leu 608
           |||
Db    675  GTA 677
```

```
RESULT 8
BM674993/c
LOCUS       BM674993             708 bp    mRNA      linear    EST   27-FEB-2002
DEFINITION  UI-E-EJ0-ahn-k-22-0-UI s1 UI-E-EJ0 Homo sapiens cDNA clone
            UI-E-EJ0-ahn-k-22-0-UI 3', mRNA sequence.
ACCESSION   BM674993
VERSION     BM674993.1    GI:18984891
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 708)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PUBMED      8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@ulowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..708
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EJ0-ahn-k-22-0-UI"
                     /tissue_type="fetal eyes, lens, eye anterior segment,
                     optic nerve, retina, Retina Foveal and Macular, RPE and
                     choroid"
                     /dev_stage="fetal and adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-EJ0"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-EJ0 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tags for this library are: fetal eyes, AGAATCAAGA
                     ; lens, CGATTAGGCA; eye anterior segment, AAATGCCGCAT;
                     optic nerve, CCAATTAAGTG; retina, CCGCG; Retina Foveal and
                     Macular, GTCC; RPE and Choroid, ACCTA. This library was
                     created for the program, Gene Discovery in the Visual
                     System, supported by National Eye Institute (NEI).
                     TAG_LIB=UI-E-EJ0
                     TAG_TISSUE=Foveal and Macular Retina
                     TAG_SEQ=GTCC"
BASE COUNT    149 a    147 c    166 g    245 t    1 others
ORIGIN
```

```
US-09-494-297-2 (1-757) x BM674993 (1-708)

Alignment Scores:
Pred. No.:                               Length:       708
Score:                   0.0533          Matches:       61
Percent Similarity:      115.50          Conservative:  38
Best Local Similarity:   39.13%          Mismatches:    77
Query Match:             24.11%          Indels:        77
DB:                      2.938           Gaps:          16
                         12
```

```
Qy    116  ProLeuGlySerAspSerSerValGlyLysTrpTyrLysHisAspGlyIleSerThr 135
           ||:|||      |||  :::    :::||    ||::|    ::|||:
Db    708  CCCACAGGAAATGGAAGGG----AAGCGGTGGCCTGAAAGGCTCGAGGGAGTCCTGACG 655

Qy    136  LysPheGluAspTyrAlaMetSerProArgIleThrGlyAspGluLeuAsnGlnLysLeu 155
           |||||    ||:                  :::          |||||||
Db    654  -----GAGGAGCAGGATGGGCAAGGCCC---------GATGAGAAGAGANGAAAGTTA 613

Qy    156  ArgAlaValMetTyrAsnGlyHisProGlnAsnAlaAsnGly----------------IleMet 171
           :::|||  |||                  :::              ||||||       |||
Db    612  AGGATGCTGGAGCAGAAC--------------------------AATGGATTCTCTTTTCTTTCATG 571

Qy    172  GluGlyLeu------------------------------GluProLeuAsnAlaIleArgVal 182
           ||:|||                                  :::          |||
Db    570  CAAGGGATCAATGGGAAAACACAAATTCAGCAACCACCTCACTCAACTCATTGACTCCGCCAACATC 511

Qy    183  ThrGlnGluAlaValTrpPtyrTyrSerAspAsnAlaProIleSerAsnProAspGluSer 202
           |||||    :::              :::||:||          |||
Db    510  CGTCAGGAGGAGGATGCC------TTTGATACAACAGTGACAATTGCT-----GAAGATGGT 463

Qy    203  PheLysArgGluSerGluSerAsnLeuValSerThrSerGlnInLeuSerLeuMetArgGln 222
           |||  :::          |||    :::    |||          :::    |||
Db    462  GGCCAGACACCATATGAAGCTACTTGCACCAAGGCTTTCAGTACCACCAGAA 403

Qy    223  AlaLeuLysGlnLeuIleAspProAsnLeuAlaThrLysMetProLysGlnValProAsp 242
           |||          :::                  |||    |||    |||
Db    402  GATCTT-----------------CCTCCACTCACAAATGGGTATCCAACATCATCAAATC---- 364

Qy    243  AspPheGlnLeuSerIlePheGluSerGluAspLysGlyAspLysTyrArgAsnLysGlyTyr 262
           |||              :::|||          ||:||::|||||      ||||:|:|
Db    363  A------------AGTGTGTGTAGAAACTCAAACCAAATACCAGTCATATAATCAG---TAT 319

Qy    263  GlnAsnLeuLeuSerGlyGlyLeu------------------ 270
           |||              |||
Db    318  CCTAATGGGTGCAGCCAATGGCTTTGGTGCAGTTAGAAAACTTTAGCCCCACTGACTATTAT 259

Qy    271  ------ValPro---ThrLysPro---------------------- 280
           ::|||     |||::|||              |||::|:||
Db    258  CATTCAGAAATTCCAAACACAAGACCACATGAAAATTCTGGAAAAAACCTTCCCTCCACAG 199
```

RESULT 9

```
Qy    281  ProProMetProProAsnGlnProThrThrSerValLeuIleArgLysTyrAlaIle 300
           ||||||   :||||      ||||||      |||:::
Db    198  CCACCACCTCCTCCTTCGGTACCACCACAACT-----------GTGATTCAAAAGAACTGGC 148

Qy    301  GlyAspTyrSerLysLeuLeuGlyLysValAlaThrLeuGln----------------- 313
           ||||||   :|||||      |||
Db    147  TCACCTGAAATAAAACTATAAAAAATAACCAAAAACTATCCAGAAATGGCAGGGAATTGTTTGAG 88

Qy    314  -----LeuThrGlyAspAsnValAsnSerPheGlnAla 324
           |||   :|||
Db     87  TCTTCCCTTTGTGGGAGACCTTTTAAATGAAGTACAGGCA 49
```

RESULT 10

(unreadable continuation)

```
LOCUS       AZ681177           863 bp    DNA    linear    GSS 14-DEC-2000
DEFINITION  ENTMR45TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION   AZ681177
VERSION     AZ681177.1  GI:11818323
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 863)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 16
            High quality sequence stop: 825.
FEATURES             Location/Qualifiers
     source          1..863
                     /organism="Entamoeba histolytica"
                     /mol_type="genomic DNA"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at the
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                     method for isolate identification. Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library construction is described in detail in Smith,
                     H.O. and Venter, J.C.    (Making small insert libraries for
                     whole genome shotgun sequencing projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT    456 a    65 c    151 g    191 t
ORIGIN
```

RESULT 10

```
LOCUS       BX350683          945 bp    mRNA    linear    EST 05-MAY-2003
DEFINITION  BX350683 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
            Homo sapiens cDNA clone CS0DL008YB09 3-PRIME, mRNA sequence.
ACCESSION   BX350683
VERSION     BX350683.1  GI:30375505
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
```

```
Qy    479  LeuLysHisIleLysLysValIleGluLysGlyIleTyrArgGluTyrAlaIleGlu 498
           |||:::|        ||||||||        :::|           :::
Db    143  TTAGAGAAAATTAAAAAAAGTTATAGAAAAAGGAATATTTAGAAGTGTAGTAGGAAAAG 202

Qy    499  TyrSerGly----------LeuThrGluThrGlnLeuArgAlaAlaThrGlnLeuAlaIle 515
           |||        :::       |||
Db    203  AAAATTGGAGGAAAAGAGTGTTAGAGAAGTCAAAAAAAAGAGAATTAGCT----CAAAATAAGAAA 259

Qy    516  TyrTyrPheThrAspSerAlaGluLeuAspLysGlySerLeuLeuLysAspTyrHisGlyPhe 535
           :::|||    |||||      ||||      :::|:::|||
Db    260  CATTATGAACAACTTGAAAAAGAATTA----AAAGAAAAACAAGAAGGAATAC-----TTT 310

Qy    536  GlyAspMetAsnAspSerThrLeuValAlaValAlaLysLysIleLeuValGluTyrAlaGlnAsp 555
           ||||||||:::||||||        |||          :::|||
Db    311  GGAGAATTAAACCATCAA--ATTGGAGTAGAAAAAAAATATTTAGAAGTGTAGGAAAGA 367

Qy    556  SerAsnProProGlnLeuThrAspLeuAspPhePhePheIleProAsnAsnAsnLysTyrGln 575
           |||        :::       |||
Db    368  ATAAATAAAGAAAGA----GAAATTGAAATGTTT---- 397

Qy    576  SerLeuIleGlyThrGlnTrpHisProGluAspPheLeuValAlaSerProIleIleIleArgMetGluAsp 595
           |||        |||    :::|
Db    398  --------AACATTCAACAAGAAAATTGATGTGTACTTAAAGAACGAATTAAAATAGAGAA 448

Qy    596  LysLysGlyValIleIleProValThrHisAsnLeuThrLeuArgLysThrValThrArgGlyLeu 615
           |||          |||          :::|
Db    449  GAGAAGAAT----AGTTCAACAAGAAAAAATGGAATTAGAAAAAGAAATTTAT----496

Qy    616  AlaGlyAspArgThrLysLysAspPheHisPheGluIleGluGluLysLysAsnAsnLysGlnGlu 635
           |||||      |||    :::
Db    497  ----GATCTTAAAGAAGGAGATTCAACAGAAAATT----AATGAAAAAGAAGAT 541

Qy    636  LeuLeuSerGlnThrValLysAlaAspLysThrAspLeuAsnLeuGluGluPheLysLysAspGlyLysSerAla 655
           |||    :::|||||      |||||||
Db    542  GAAATTAAACAAAAGCTTGAAAGAAATAATTACAGAAAATAAAAGTGGAGAATTA 601

Qy    656  ThrIleAsnLeuLysHisGlyGluSerLeuThrLeuGlnGlyLeuProGluGlyValTyrSer 675
           |||||      |||    |||
Db    602  GCTACTGTATTAAAGACAAAATGGAAAAACTAAAAATGTTACAGATAAAGAGAAGA 661

Qy    676  Tyr----------LeuValLysGluThrAspSerGluGlyTyrTyrLysValLysVal 690
           |||        :::       |||
Db    662  TATGAAGAAGTTGAAAATCTTATTAAACAAAATCAAAAGAA---703

Qy    691  AsnSerGlnGluValAlaAsnAlaThrValSerLysLysThrGlyIleThrLysSerTyrAspSerGluThr 710
           |||          |||          |||
Db    704  ------CAAAGTCATCGACAAAGTATGTATAGTAAAATTGGAAGTTGGATAGATGAAACT 757

Qy    711  LeuAlaPheGluAsnAsnLysGluProValValProThrGlyValAspGlnLys 728
           |||          |||    :::|||
Db    758  GTTCGAAGTGCTAGAAATGGAACAAATAGAAATTACCTATTAAAGACAATGAAAAG 811
```

```
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2129.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAIO45ZA06_CS04265_16cluster=2129.f.
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAIO45ZA06_CS04265_1.

FEATURES             Location/Qualifiers
     source          1..945
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DL008YB09"
                     /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
                     25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT      219 a    239 g    303 t    183 c
ORIGIN
```

US-09-494-297-2 (1-757) x BX350683 (1-945)

Alignment Scores:
```
Pred. No.:                    0.182
Score:                        112.50
Percent Similarity:           33.60%
Percent Identity:             33.85%
Best Local Similarity:        23.85%
Query Match:                  2.85%
DB:                           13
```

```
                      Length:          945
                      Matches:         88
                      Conservative:    36
                      Mismatches:      145
                      Indels:          101
                      Gaps:            15
                                       1 others
```

```
Qy    164 ProGlnAsnAlaAsnGlyIleMetGluGlyLeuGlyLeuGlnProLeuAsnAlaIleArgValThr 183
          |||  :  ||    |  |||   |||   |||||  |||||     |||
Db    885 CCCCGNCCCAGTCAGGGAAAGATNAAAAGGTTTGGAATCA                              844

Qy    184 GlnGluAlaValTrpThrTyrTyrSerAspAsnAlaProIleSerAsnProAspGluSerPhe    203
          ||:||:::||:|  :  |||| |           ||||:::||:|||||||
Db    843 CAACAGTCGGTAGTGTAC------AACACAGGTATTCTTCTTGCCCAGTCTCAAGCAC         793

Qy    204 LysArgGluSerGluSerAsnLeuValSerThrSerGlnIleLeuSerLeuMetArgGlnAla    223
          |||                   |||     ||  :||:||:
Db    792 CCAATATATATCAAGGAGACAGACATGTC------CAACAGTCAATAGGTGTGGGACACATCAC   742

Qy    224 LeuLysGlnLeuIleAspProAsnLeuAlaThrIleLysMetProLysGlnVal----------  240
          |||                            |||    |||   ||||||
Db    741 CTTAT-CCACAGACAACTCCACAATTGTACAGAGTTATGCCCAGCCAAGTCTTCAGTATA      683

Qy    241 ------ProAspAspPheGlnLeuSerIlePheGluSer----------GluAspLysSerGly  255
                |||  ||    ||||                      ||  |||:
Db    682 CCAGGGGCAACACAGATTTTCACAGCTCATCCACAAGGAGTGGTGGTACACCCAGCGCAC      623

Qy    256 AspLysTyrAsnLysGlyTyrGln-----------------------------Asn-LeuLeu  266
          ||||||                                             ||| :::::
Db    622 AGTGACTACACAAGTTGCACCAGGGCAGCCGCCTCAGCCCTTGCACCCATCTGAAANGGTTGT 563

Qy    266 uSerGlyGlyLeuValProThrHisProProGlyAspProPro-------------Me      283
          ||||                        ||||||||     |||                ::
Db    562 GACAAAATAATCTCTTGGATCTGCGCCCCCTCTCCCAAACCAAAAACCATTGTCTT         503

Qy    283 trProProAsn-------------GlnProGlnThrThrSerValLeuIleArgLysTy      298
          :||:||||||||                |||||||:|||||:||
Db    502 ACCTCCCAACTGGAAGACAGCTCGAGATCCAGAAGGGAG------ATTTATTACTA        452

Qy    298 rAlaIleGlyAspTyrSerLysLeuGluGlyLeuGlnLeuThrGlyAspAs              318
          :||                  ::    :: ||:  ||
Db    451 CCATGTGAATCACAAGGACAGACTCAGTGGGTAGTCCTCTTACTTGGGAAAGCCCAGAGATGA 392

Qy    318 nValAsnSerPheGlnAlaAArgValPheSerSerAsnAspIleGlyGluArgIleGlu      338
          :  :::                           |||::
```

```
FEATURES          Location/Qualifiers
     source       1..2421
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:3050255"
                  /tissue_type="Skin, melanotic melanoma."
                  /clone_lib="NIH_MGC_20"
                  /lab_host="DH10B-R"
                  /note="Vector: pOTB7"
```

```
BASE COUNT      661 a      642 c      692 g      426 t
ORIGIN
```

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 22 Row: g Column: 8
This clone has the following problem: frame shifted.

Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

```
US-09-494-297-2 (1-757) x BC008707 (1-2421)

Alignment Scores:
Pred. No.:       1
Score:           112.00
Percent Similarity: 33.41%
Best Local Similarity: 20.73%
Query Match:     2.84%
DB:              11

                    Length:        2421
                    Matches:       85
                    Conservative:  52
                    Mismatches:    141
                    Indels:        132
                    Gaps:          20
```

```
Qy   67  SerSerGluTyrArgTrpTyrGlyTyrGluSerTyrValArgGlyHisProTyrTyrLys  86
             ||||||   :::|||  |||   ||||||::: ||:    ||||
Db  849  AACTCCCAGTTTGACTTTCTGCGGTTCGACACATGAAGAT-----AACCCCTACTATAAG  902

Qy   87  GlnPheArgValAlaHisAspLeuArgValAlaAsnGlyLysSerArgSerTyrGlnVal  106
             :::||||    |||       ||||||     |||
Db  903  TTCATCCAGAAAGCCATGAAA-----------------------GAGGGA-----CGGTACACTGTC  941

Qy  107  TyrCysPheAsnLeuLysLysAlaPheProLeuGlySerAspSerSerValLysLysTrp  126
             |||
Db  942  CTGGCAGAAAAC----------------------------------------AAAAGTGAC  962

Qy  127  TyrArgLysHisAspGlyIleSerThrHisLysPheGluAspTyrAlaMetSerProArgIle  146
             ||||||   :::    ||::||  ::||||||||||||   ::
Db  963  GAGAAAAAAAATCAGGAGTGCAGCGTCTGACAATGAAGAT-----GATGATGATGAAGAA  1016

Qy  147  ThrGlyAspGlyLeuAsnGlnIleLeuArgAlaValMetTyrAsnGlyHisProGlnAsn  166
             ||:||
Db  1017  GATGGGAAATTACCTTCATCCCTCTCCTTTGCC----------------TCCAAGAAG  1058

Qy  167  AlaAsnGlyIleMetGluGlyLeuGluProLeuAsnAlaIle-------------------  180
             ||| :::  ||||:::|||:|:||||||   :::
Db  1059  TGTAACCGCCTTGAAGAGCTGAATGAGAGCCCTTGAAGTAGTGGACCCCAGATCATCCCCTC  1118

Qy  181  ---ArgValThrGlnGluAlaValAlaTrpTyrTyrSerAspAsnAlaProIleSerAsnPro  199
             ::::::|||::|||::||||||
Db  1119  GCAGGCACTTGTTCGTAAGGCACAGCGTGACAGTTCCACTCCACCCCA-----CACAACGCA  1175

Qy  200  AspGluSerPheLysArgGluSerGluSerAsnLeuValSerThrSerGlnLeuLeuSerLeu  219
             |||     |||   |||:|||   |||
Db  1176  GACGGTGCCGCTGTGCAGCCCTCCAGGTGGAGTACACGCGCAGACTCGACCGTTGGCAGCC  1235

Qy  220  MetArgGlnAlaLeuLysGlnLeuGlnLeuAspProAsnLeuAlaThrLysMetProLysGln  239
             ||| ||   :::  |||
Db  1236  ATGTRTTACAGCTACTACACATGCTACCGGACGGCACTYACTYACGCGTGGCGCGCCCC  1289

Qy  240  ValProAspAspPheGlnLeuSerIlePheGluSerGlnAspLysGlyAspLysTyrAsn  259
             ||| |||
Db  1290  ---CCTCCCGGAAATCGGATGTGAACTACTTAC----------------------------  1316

Qy  260  LysGlyTyrGlnLeuAsnLeuSerGly-------------------------GlyLeuValPro  272
             |||:::|||   |||::|||
Db  1317  ------TACAGCACCCTTCCTCCTGGGGTGGGTGACCGTGTGTCTAACTCCCCTGGGAGTGACGACC  1370

Qy  273  ThrLysProProThrProGly---AspProProMetProProAsnGlnProGlnThrThr  291
             |||||||    |||||||   :::|||:::||  ||||||::: |||||:::
Db  1371  ACCGGCCCCACCACCTCCTTGGGACCACACCACACCGGCCCCCAACCACAGCAGAACTTAGC  1430

Qy  292  SerValLeuIleArgLysTyrArgAlaIleGlyAlaAspTyrSerLysLeuLeuGlyTyrAlaThr  311
             ::|||||||   |||    ||||||   ::|||||:::||:|||:||
Db  1431  AGC-----------------------------------------------GGGGCCACC  1442

Qy  312  LeuGlnLeuThrGlyAspAsnValAsnSerPheGlnAlaArgValPheSerSerAsnAsp  331
             |||
Db  1442  ------------------------------------------------------------  1442

Qy  332  IleGlyGluArgIleGluLeuSerAspGlyThrThrTyrThrLeuThrLeuGluAsnSerPro  351
             ::: |||:::|||
Db  1443  ------------------------------------------TCCACAACACACCACCACACAAGCTGCACTTGGCCCC  1475

Qy  352  AlaGlyTyrSerIleAlaGluProIleThrProLysLysValGlnGluAlaGlnLysLysVal------  369
             ||| ::: |||
Db  1476  GTGGCCCGGCCATCATCCCCCCGCCCCCCGACGTCCAGCCCGTGGATTGACAAGGCTGGGCCGCCAG  1535

Qy  370  TyrThrIleIleAspGlyLysGlyGlnIleIleGluAlaAspAsnProAsnLys---------  383
             ||| ||| ::: |||:::::  :::::
Db  1536  TATGTCGCCAGGAAACGGCCCTGAAGTTCGGTCCAGAGTTCCGTCCCCAGAGAATGATCAAAGA  1595

Qy  384  ---GluIleValGluLeuProTyrSerValGluValAlaTyrAsnAspPheGlyGluPheSerVal  402
             ||| ::: |||||||
Db  1596  TTTGAGTTCCTGCAGCCGTGG---------------CACCAGTATAAATGCTTATTATGAGTTTAAG  1646

Qy  403  LeuThrThrGlnAsnTyrAlaLeuLysPheTyrTyrTyrAlaLeuLysAsnLysAsnGlySerSerGln  422
             |||::|||   |||||||   |||:::|||
Db  1647  ------------------AAGCAGTTGTTCTTCCTGCCAGAAAAGAGGGGGGATAGCACATGCAG  1688

Qy  423  ValValTyrCysPheAsnAlaAspLeuLysSerProProAspAspSer----------------  437
             :::
Db  1689  GCTGTG------------TCTGCACCAGAGAGAGGGCTCCCACAGACTTCTGTCCCCCAGAGAACCCA  1739

Qy  438  GluAspGlyGlyLysThrMetThrProAsp  447
             |||:::|||
Db  1740  ACTGATGCTGGGGGGAGGAATGGCGCGCCCTGA  1769
```

```
RESULT 12
BC015953
LOCUS       BC015953                2430 bp    mRNA            linear    HTC 24-OCT-2001
DEFINITION  Homo sapiens, Similar to splicing factor, arginine/serine-rich 8
            (suppressor-of-white-apricot, Drosophila homolog), clone
            IMAGE:3830054, mRNA.
ACCESSION   BC015953
VERSION     BC015953.1   GI:16358985
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2430)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan
```

```
FEATURES        Clone distribution: MGC clone distribution information can be found
                through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
                Series: IRAL Plate: 14 Row: j Column: 1
                This clone was selected for full length sequencing because it
                passed the following selection criteria: Hexamer frequency ORF
                analysis. Similarity but not identity to protein
                This clone has the following problem: frame shifted.
source
     1..2430
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="IMAGE:3830054"
     /tissue_type="skin, melanotic melanoma."
     /clone_lib="NIH_MGC_20"
     /lab_host="DH10B-R"
     /note="vector: pOTB7"
BASE COUNT    663 a    644 c    697 g    426 t
ORIGIN
```

Alignment Scores:
```
Pred. No.:                      0.921
Score:                        112.00
Percent Similarity:            33.41%
Best Local Similarity:         20.73%
Query Match:                    2.84%
DB:                            11
```

```
Alignment No.:  1
Length:         2430
Matches:        85
Conservative:   52
Mismatches:     141
Indels:         132
Gaps:           20
```

US-09-494-297-2 (1-757) x BC015953 (1-2430)

```
Qy     67 SerSerGluTyrArgGlyIleSerThrLysPheGluSerTyrValArgGlyHisProThrTyrLys 86
          ||||||    ||||    |||::||:|||||  |||||        ||||||:||||||||||||
Db    858 AACTCCCAGTTTGACTTTCTCCGCTTCGACCACTACCTC----AACCCCTACTATAAG 911

Qy     87 GlnPheArgValAlaHisAspLeuArgValAlaValAsnLeuGluLysGlySerTyrArgGlnVal 106
          |||:||    ||||||    |||||||    | || |||||        |||  |||||||||
Db    912 TTCATCCAGAAAGCCATGAAA---------------GAGGGA----CGCTACACTGTC 950

Qy    107 TyrCysPheAsnLeuLysLysAlaPheProLeuGlySerAspSerSerValLysTrp 126
          |||                    ||| |   |  |      |||||||||||||||
Db    951 CTGGCACAAAAC-------------------------AAAAGTGAC 971

Qy    127 TyrLysLysHisAspGlyIleSerThrLysPheGluSerTyrAlaMetSerProArgIle 146
          |||||||   ||||          ||::|||||||||||      ||||||||||||
Db    972 GAGAAAAAAAATCAGGAGGTCAGCTCTGACAATGAAGAT----GATGATGATGAAGAA 1025

Qy    147 ThrGlyAspGlyLeuAsnGlnIleLysGlyLeuArgGlyAlaValMetTyrAsnGlyHisProGlnAsn 166
          |||||:||    |||||:||||    |    |||||          ||:
Db   1026 GATGGGAATTACCTTCATCCCTCTCTCTTGCC---------------TCCAAGAAG 1067

Qy    167 AlaAlaAsnGlyIleMetCysGlyLeuGluProLeuAsnAlaIle------------- 180
          ||:         ||| |||  |  |  | |||| |||||||||||
Db   1068 TGTAACCGCCTTGAAGAGCTGATGAAGCCCTTGAAGGTAGTGGACCCAGATCATCCCTC 1127

Qy    181 --ArgValThrGlnGluAlaValTrpProTyrTyrSerAspAsnAlaProIleSerAsnPro 199
          |||:|     |||||||||||||||||||||||||          ||| |
Db   1128 GCAGCACTGTTCGTTGGAAGGCACAGCGTGACAGTTCCACCCACCCA---CACAACGCA 1184

Qy    200 AspGlyLeuSerPheLysArgGlyUserGlnSerGlnSerGlnIleLeuSerLeu 219
          |||||||:   ||| |  || |                  :|| |   |  |
Db   1185 GACGGTGCGCCTGTCGCAGCCCCTCCAGGTGAGCTGAGTACAGCAGGCTGCCAGCC 1244

Qy    220 MetArgGlnAlaLeuProLeuAspProAsnLeuAlaThrLysMetProLysGln 239
          |||||||||||||   |||| |   |||||||||||          |
Db   1245 ATGTAYTACAGCTACTACACATGCTACCGGACGCACTTACTGCCTGGGCGCCGGCC---- 1298

Qy    240 ValProAspAspPheGlnLeuSerIlePheGlySerGluAspGlyAspLysTyrAsn 259
          |||       |||           |::|||          :|||
Db   1299 --CCTCCCGGAATCAGATGTGACTACTTAC---------------GGAGTAGTACTTAC 1325

Qy    260 LysGlyTyrGlnAsnLeuLeuSerGly------------------GlyLeuValPro 272
          |||          |||  ||                    |||:::|||
Db   1326 ------------------TACAGCACCCTTCCTGCTGCGGGTGACCGTGTCTAACTCCCCTGGAGTGAGCGACC 1379
```

```
Db   1380 ACCGCCCCACCACCTCTGGGACCACCACCAGCCCCCAACCACACCAGCAGACTAGC 1439

Qy    292 SerValLeuIleArgLysTyrAlaIleGlyAspTyrSerLysLysLeuLeuGlyGlyAlaThr 311
          ||| |||||||   ||                  ||||           |||||||:
Db   1440 AGC----------------------------------GGGGCCACC 1451

Qy    312 LeuGlnLeuThrGlyAspAsnValAsnSerProAsnAlaArgValPheSerSerAsnAsp 331
          |||||||||||||||                          |||||:
Db   1451 GTGGCCGCCACCATCACCCCCCCGCCCCCCCCGACGTCCAGCCGCGTGGAGCGCCGAG 1544

Qy    332 IleGlyGluArgIleGluLeuSerAspGlyThrThrTyrThrLeuThrLeuAsnSerPro 351
          ||||:||||||||||||:  |          |||            |:    |
Db   1545 TAYGTCGGCAGGAAACGAGCGCCTGAAGTTCGAGCACACCAGTGTTGTGCCAAGAATGATCAAAGA 1604

Qy    352 AlaGlyTyrSerIleAlaGluProIleThrPheLysValGluAlaLeuGlyLysValGlyVal 369
          |||          |||||||||||||||||||||          |||||::
Db   1605 TTTGAGTTCCTGCAGGCCTGG-----CACCAGTATAATGCTTATTATGAGTTTAAG--- 1655

Qy    384 ----GluIleValGlyProTyrSerValGluAlaTyrAsnAspPheGluGluPheSerVal 402
          ||| |||  ||  |  ||| |||||||| ||::
Db   1698 GCTGTG---------------TCTCGCACCAGAAGAGGCTCCCCACCACAGACTGCTCCCGGAAGCCA 1748

Qy    438 GluAspGlyGlyLysThrMetThrProAsp 447
          |||||||  | |  ||||:||
Db   1749 AGTGGATGGTGGGGAGGAGGATGGCGCGCCTGAA 1778
```

RESULT 13
AK076636
```
LOCUS       AK076636               1493 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male testis cDNA, RIKEN full-length enriched
            library, clone:4922502J23 product:similar to RSD-6 [Rattus
            norvegicus], full insert sequence.
ACCESSION   AK076636
VERSION     AK076636.1  GI:26345549
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
```

```
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
    TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
   MEDLINE   21085560
    PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
    TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1493)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,N., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
    TITLE   Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9216,
            Fax:81-45-503-9216)
  COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..1493
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
```

REFERENCE (top)
```
  AUTHORS   Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,K., and Hayashizaki,Y.
    TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   MEDLINE   20530913
    PUBMED   11076861
```

```
     CDS             /db_xref="FANTOM_DB:4922502J23"
                     /db_xref="taxon:10090"
                     /clone="4922502J23"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
                     43..1218
                     /note="unnamed protein product; putative"
                     similar to RSD-6 [Rattus norvegicus] (SPTR|O9J1I6,
                     evidence: FASTY, 74.9%ID, 99.6%length, match=762)"
                     /codon_start=1
                     /protein_id="BAC36426.1"
                     /db_xref="GI:26345550"
                     /translation="MAEDGSPKIYSRPPRDNSKTPTEADIFFGADNTIPKSETTITSE
                     GDHVTSVNDCTPDGDFSTTVNKLFPTKEKLKIEDDIEGCLKLTTIPTPTETPN
                     SKPKGSITENFIPVKIGNTSSPVGTVSLLIDFSSNTAKEDIFLTTIDTGEKEVPTTEP
                     SGTLEDSAADVEDASGFPDESTETDVPSSATSDAPDDGAVQVTDSFSPEAGVPPSTEK
                     EVTIPIDTNIAEENVTEIKLIVSEDRPKTVTKLSDSDSEEEKFIVFELNSAEKAKDN
                     VEDPLNDEESTDGANDWMEKETASEAESHAVLITAVESRYDFIVTASETDNVMEESHV
                     NTTDLPENETESVTNVTEELPSVTSIVDTLKDKEDLSTTNSGLFKLKEEPDDLMM"
     polyA_signal    1470..1475
                     /note="putative"
     polyA_site      1493
                     /note="putative"
BASE COUNT      493 a    326 c    305 g    369 t
ORIGIN
```

```
US-09-494-297-2 (1-757) x AK076636 (1-1493)

Alignment Scores:
 Pred. No.:          0.484              Length:        1493
 Score:            111.50              Matches:         97
 Percent Similarity:  34.54%           Conservative:    46
 Best Local Similarity: 23.43%         Mismatches:     156
 Query Match:         2.83%            Indels:         115
 DB:                      11           Gaps:            21
```

```
Qy   311 ThrLeuGlnLeuThrGlyAspAsnValAsnSerPheGlnAlaArgValPheSerSerAsn 330
         |||::|
Db   160 ACTATTACCTCGGAAGGAGAccACGTTACTTCAGTA--------------------- AAT 198

Qy   331 AspIleGlyGluArgIleGluLeuSerAspGlyThrThrThrLeuThrGluAsnSer 350
         |||||||               ||:|:::||            |||
Db   199 GACTGC----------------- ACACCGGAtGGTGATTTTCAACTACA---------- 231

Qy   351 ProAlaGlyTyrSerIleAlaLeuProIleThrPheLysValGluAlaGlyLysValTyr 370
         |::||            |||||||                     |||:::||
Db   232 ----------------- GTCAACAAGCTCACACCTTACAAAGGAAAAAACTCAAAATAGAA 273

Qy   371 ThrIleIleAspGly-----LysGlnIleGluAsnProAsnLysGluIleValGluPro 388
         ||:|:||||:||          |||             |||||||
Db   274 GATAGATATGTTGAAGGCTGTCTGAAAATTAACAAACTCTTCCAGAGAAGAAAATTACCACTCCA 333

Qy   389 TyrSerValGluAlaTyrAsnAspPheGluPheSerValLeuThrThrGlnAsnTyr 408
         |||                      :||                ||:|:::|||
Db   334 ----- ACTGAAAACTCCAAACTCCAAAACCTAAGGGGTCAATT------ ACTGAAAAT--- 378

Qy   409 AlaLysPheTyrTyrAlaLysAsnGlySerSerGlnValValTyrCysPheAsn 428
         |||             |||||                     |||
Db   379 ----- TTCATTCAGTGAAAATTGGAAATACCTCATCCCAGTT---------------- 417

Qy   429 AlaAspLeuLysSerProProAspSerGluAspGlyGlyLysThrMetThrProAspPhe 448
         |||                 :::                    ||:|:::|||
Db   418 ----------------------- GGTACTGTTTTCTTTAATAGATTTT------------ 441

Qy   449 ThrThrGlyGluValLysTyrThrHisIleAlaGlyArgAspLeuPheLysTyrThrVal 468
         |||:::||                   ||||:|::|||
Db   442 TCCAGTAACACGGCAAAA---------------------- GAAGATATCTTCTTAACCACC--- 480

Qy   469 LysProArgAspThrAspProAspThrPheLeuLysHisIleLeuLysLysValIleGluLys 488
         |||:::|                      ||||:::
Db   481 ------------------------------------------- ATTGACACA 489
```

```
Qy    489 GlyTyrArgGluLysGlyGlnAlaIleGluTyrSerGlyLeuThrGluThrGlnLeuArg 508
              |||           :::||:||
Db    490 GGAGAAAAAGAGAGTCGTACCACTACTGAGTTCTCTGGCACACTAGAG----------- 537

Qy    509 AlaAlaThrGlnLeuAlaIleTyrTyrPheThrAspSerAlaGluLeuAspLysAspLys 528
              ||||:::::::||||||||
Db    538 ------------------------------------GACAGTGCTCCT---------GAT 552

Qy    529 LeuLysAspTyrHisGlyPheGlyAspMetAsnAspSerThrLeuAlaValAlaLysIle 548
              |||||::||                     |:|
Db    553 GTGGAGGATGCCTCTGGATTTCCAGATGAAAGCACTGAAACT--GANGTTCCTTCCTCT 609

Qy    549 LeuValGluTyrAlaGlnAspSerAsnProProGlnLeuThrAspLeuAspPhePhePhe 568
                          |||||:::||||              |||
Db    610 GCCACCTCAGATGCTCCTGACTACATGGAGCTGTTTCAGGTCACT------GACTCCTTTAGT 663

Qy    569 ProAsnAsnAsnLysTyrGlnSerLeuIleGlyThrGlnTrpHisProGlnAspLeuVal 588
              |||            |||||||||:
Db    664 CCTGAGGGCTGGAGTCCCCCCCTCA------------ACTGAAAAGGAGGTCCACCACCATTCCA 714

Qy    589 AspIleLeuArgMet---GluAspSerGlyValValIleProValThrHisAsnLeuThr 607
              |::||         |:::                   :::
Db    715 GACATAACCAGAAGAGAAGAGTAACTGAGATCAAATTGATTGTTTCAGAAGAT 774

Qy    608 LeuArgGlyThrValThrGlyLeuLeuAlaGlyAspThrLysAspSerPheHisGluIle 627
              ||||||||||||||          |||:
Db    775 AGGCCCAAAACTGTGACTAAACTAAGTGACTCCGGAGAGAAAAATTTATCACTGTTTTTT 834

Qy    628 GluLeu------------------------LysAsnAsnLysGlnGluLeuSerGlnThr 640
              ||||||                        |::|:||
Db    835 GAACTCACTAACTCGGCTGAAAAAAGCCAAAGATAATGGGGAAGATCCTTTAAATGATGAG 894

Qy    641 ValLysThrArgAspLysThrAsn--------LeuGluPheLysAspGlyLysAlaLysIleAsn 658
              |||||||||                           :::
Db    895 GAATCGACTGATGGAGCTAATGATTGGATGGAA-------AAAGAAACTGCAAGT 942

Qy    659 LeuLysHisGlyGluSerLeuThrLeuGlnGlyLeuProGlnGlyTyrSerTyrLeuVal 678
              :::|||                |||          :::::::|||
Db    943 GAGGCAGAGTCCCATGCTGTCTTCTGCACTAATGCAGGTATGACNTTANTGTC 1002

Qy    679 -----LysGluThrAspSerGluGlyTyrLysValLysAsnSerGlnLeuValAla 696
                   |||||||:|||::
Db    1003 ACTGCCTCGAAAACTGACAACGTCATGGAAGGAAGTCACAGTGAACACACCACAGATTTGCCT 1062

Qy    697 -----AsnAlaThrValSerLysThrGlyIleThrSerAsp 708
                   |||     |||          :::
Db    1063 GAAAACAGACAACAAGAGTCTGTCCACTAATGTCACAGAGGAG 1104
```

RESULT 14
```

FEATURES
    URL:http://genome.gsc.riken.go.jp/
    URL:http://fantom.gsc.riken.go.jp/.

    source   Location/Qualifiers
                    1..2167
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /mol_type="mRNA"
                    /db_xref="FANTOM_DB:6030440B06"
                    /db_xref="taxon:10090"
                    /clone="6030440B06"
                    /sex="male"
                    /tissue_type="testis"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="13 days embryo"

    misc_feature   37..1633
                    /note="chaperonin subunit 6a (zeta) (MGD|MGI:107943,
                    GB|NM_009838, evidence: BLASTN, 99%, match=2161)
                    putative"
    polyA_signal   2150..2155
                    /note="putative"
    polyA_site   2167
                    /note="putative"

BASE COUNT   630 a   430 c   549 g   558 t
ORIGIN

US-09-494-297-2 (1-757) x AK077916 (1-2167)

Alignment Scores:
Pred. No.:
Score: 0.873
                  111.50
Percent Similarity: 33.33%
Best Local Similarity: 20.99%
Query Match: 2.838
DB: 11

Length: 2167
Matches: 102
Conservative: 60
Mismatches: 169
Indels: 155
Gaps: 20

Qy 146 IleThrGlyAspGlnLeuAsnGlnLysLeu--------------------------ArgAla 157
        ||||||| ::||
Db 295 ATAACTGGCGATGGCACTACATCCAATGTCTCATCATCGGGGAGCTGCTCAAAACAGGCG 354

Qy 158 ValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyIleMetGlyLeuGluProLeu 177
        :::||| ||| |||||||||| |||
Db 355 GACTGTAC-----------------------ATTTCTGAAGGTCTTCACCCA--- 384

Qy 178 AsnAlaIleLeuArgValThrGlnLeuAlaValThrpIleTyrTyrSerAspAsnAla--ProIle 197
        |||:::: ||||||||||||| ||| ::| ||||||||
Db 385 ------------AGAATAATAACTGAAGGTTTTGAAGGCGGCAAAAGAAAAAGGCACTCCAAATTT 435

Qy 197 erAsnProAspGluSerPheLysArgGluSerAsn-LeuValSerThrSerGln 216
        :::||||| ||||| |||||| |||||||
Db 436 CTGGAACAACGTCAAAGTAACCAAAACAGGATGGACAAGAAACACTCAInCGAnGTGGCGAGG 495

Qy 217 LeuSerLeuMetArgGlnAlaLeuLeuLysGlnLeu------------------------Ile 228
        ||||||| :::
Db 496 ACAICTCTGCGGACTAAAGTTCAInGCTGACTGAACTTGCAGAGTGTCTTGACAGAGGCTGTAGTG 555

Qy 229 AspProAsnLeuAlaThrLysMetProLysGlnValProAspAspAsnPheGlnLeuSerIle 248
        ||| ||| ||| ||| :::
Db 556 GACTCCCATCTTGGCCATTAGGAAAAAGAACGACGAGCCCCATTGACCTCTTCATGGTTGAGATC 615

Qy 249 PheGluSerGluAspLysGlyAspLysTyrAsnLysGlyThrGlnAsnLeuLeuSerGly 268
        ||| ::: |||
Db 616 ATGGAGAGTGAAGCAATAAAATCT----------------------------- 636

Qy 269 GlyLeuValProThrLysProProThrProGlyAspProMetProProAsnGlnPro 288
        ||| :::
Db 636 -- 636

Qy 289 GlnThrThrSerValLeuArgLysTyrAlaIleGly----------------- 302
        :::||| ||||||||||||| :::
Db 637 GAGACAGATACAAACGTTAATCAGAGAGGGCTTGTTTTTGGATCATGGAGCTCGGCATCCTGAT 696

Qy 303 TyrSerLysLeuGlnLeuGluGlyVal------------------ThrLeuGlnLeuThr 315
        ||| :::||| :::|||
Db 697 ATGAAGAGAGAGACTGGAAAATGCCTACATCCTCACTTGCAACGTGTCCTTAGAGTATGAG 756

Qy 316 GlyAspAsnValAsnSer---------------PheGlnAlaArgValPheSerSerAsnAsp 331
        ||||||| ::: ||| |||
Db 757 AAAACAGAACGAAGTGAATTCTCGGGTTTTTTTTTTACAAGAGTGCCAAGAGAGAGAAAAACTAGT 816

Qy 332 IleGly-GluArgIleGluLeuSerAspGlyThrTyrThrLeuThrGluLeuAsnSerPr 351
        ||| |||||||| ::: ||| ||||||||||||||||||||||
Db 817 AAAGGCTGAAAGAAAAATTCAInTGAAGATAGAGTTAAAAAAAATCAInTAGAGCTGAAAAAG-- 874

Qy 351 oAlaGlyTyrSerIleAlaLeuGluProIleThrPheLysVal---------GluAlaGlyLysVa 369
        ||| ||| |||||| :::
Db 875 -------------------------------AAAGTCTGTGGTGACTCAGATAAAGG 900

Qy 369 lTyrThrLeuGlnIleLysGlnIleIleGluAsnProAsnLysGluIleLeuLevalGluProTy 389
        ||||||| ::: |||||||||
Db 901 ATTTGTCGTTAInT-------------------------AInCAAAAGGGATTGACCCCTT 936

Qy 389 rSerValGluAlaThrTyrAsnAspPheGluLeuPheSerValLeuThrThrPheGlnAsnTyrAl 409
        :::|||||||||| |||| |||
Db 937 TTCCTTAGAInTGCC-------------------------------CTTGCGAAAGAAGGAInTGT 969

Qy 409 aLysPheThrTyrThrAlaLysAsnLysAsnGlySerSerGlnValValTyrCysPheAsnAl 429
        ||||||| :::||||||
Db 970 AGCTCTCGCGCAGAGCCAAGAGAGAAACAInGAGAGGCGAGGAAGGTTCACCTTTAInTGA 1029

Qy 429 aAspLeuLysSerProProAspSerGluAspGlyGlyLysThrMetThrProAsp------Ph 448
        ||| ||| |||||||||||
Db 1030 AGCTCTGInTCGInTCCInTTGAInGAC------------CTGAAInCCTGACTGInTT 1068

Qy 448 eThrThrGlyGluValLysThrTyrThrHisIleLeuAlaGlyArgAspLeuPheLysTyrThrVa 468
        ||| ||| |||
Db 1069 GGGACAInCCAGAGCTTGInCTAInGAGInTATAInCACTGGInTGInGGAInGGAInTTCAInGA 1128

Qy 468 lLysProArgAspThrAspProAspThrPheLeu-----------------------LysHis--- 481
        ||| ::: |||||
Db 1129 GAAGTGTAInCAAInTCCInCCGInTTCInTGCInCInTTAInCTGGInTGGInGGInCCAAAInTAAInGACInCACT 1188

Qy 482 -------IleLysLysValIleGlyLysGlyValTyrArg------------------GluLy 493
        ||| ||| :::
Db 1189 GAInCTCAAInTCAInCAInGGInTGCInTInTAInGInGInTGGInCTTGInGGInCTGInTCInAAInAInTGInCInInTInTGInTGInA 1248

Qy 493 sGlyGlnAlaIleGluTyrSerGlyLeuThrGluThrGlnLeuArgAlaAla-------- 510
        ||| :::|||| :::|||
Db 1249 TGGCTGInTGInTGInTCCInCGGInGGInTGCInGGInTGCInGInTGGInGInGGInCInGGInCCInCGGInGGInCTTCInGInATTInA 1308

Qy 511 -----------------ThrGlnLeuAlaIleTyrTyrPheThrAspSerAl 522
        ||| ||| |||
Db 1309 AInTACInAInCCCInCInGTGTGInAInGGGInGGInGGGInGGInCGGInCCInGGInCGInTTGInGInGInTGInCInGInTInTGInCInGInTGGInCCInTT 1368

Qy 522 aGluLeuAspLysValAspLysLeuLysAspTyrHisGlyPheGluGlyMetAsnAspSerTh 542
        ::: ||| ||| |||
Db 1369 GCTCInTCInAInTTCCInCInAInGGInTTCInTGCGInCInAInAInInCTInGGGInTInTT---GInCInCInTTCInAInGInGInAInInCInTT 1425

Qy 542 rLeuAlaValAlaLysIleLeuValGlnTyrAlaGlnAspSerAsnProProGlnLeuTh 562
        ||| |||||
Db 1426 AGInTInTInAAInGInTInT-------------------------CInAInInGInGInTGInCInGInCInATTCInGGInGInInTCInCGGInCCInGGCInTGInT 1467

Qy 562 rAspLeuAsp 565
        :::|||
Db 1468 AGGInTGTGGInT 1477

RESULT 15
BH132924

LOCUS   BH132924   889 bp   DNA   linear   GSS 07-AUG-2001
DEFINITION   ENTMX37TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
             genomic, genomic survey sequence.
ACCESSION   BH132924
VERSION   BH132924.1   GI:15091985
KEYWORDS   GSS.
SOURCE   Entamoeba histolytica
  ORGANISM   Entamoeba histolytica
             Eukaryota; Entamoebidae; Entamoeba.

```
REFERENCE    1 (bases 1 to 889)
AUTHORS      Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE        Determination of clone end sequences from Entamoeba histolytica
             HM1:IMSS sheared DNA library (2001)
JOURNAL      Unpublished
COMMENT      Contact: Brendan J Loftus
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0208
             Fax: 301 838 3543
             Email: bjloftus@tigr.org
             Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
             DNA library
             Seq primer: M13-Forward
             Class: shotgun
             High quality sequence start: 18
             High quality sequence stop: 811.
FEATURES             Location/Qualifiers
     source          1..889
                     /organism="Entamoeba histolytica"
                     /mol_type="genomic DNA"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                     method for isolate identification. Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + 1 method used for
                     the library construction is described in detail in Smith,
                     H.O. and Venter, J.C. (Making small insert libraries for
                     whole genome shotgun sequencing projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT    399 a    72 c   154 g   264 t
ORIGIN
```

```
US-09-494-297-2 (1-757) x BH132924 (1-889)

Alignment Scores:
Pred. No.:
Score:                   111.00        Length:           889
Percent Similarity:      38.24%        Matches:           63
Best Local Similarity:   20.59%        Conservative:      54
Query Match:             2.81%         Mismatches:        87
DB:                      28            Indels:           102
                                       Gaps:              15
```

```
Db      379 TTCATAAATCAACAAGAATCAAATTAACAGAAATTAACGAAAATGAAAATTAACAAACTGTAAAAAATACGGAA 438
Qy      581 ---GlnTrpHisProGluAspLeuValAspIleIleLeuArgMetGluAspLysLysGluVal 599
                :::       :::||||||| :   :::||||:||   :::||||:::
Db      439 ATTAAAAGGAGAGTTAGATAATAATATTGAGAAAATCATCACAATAGAATAAATGAAGAAATGTAAT 498
Qy      487 GluLysGlyTyrArgGlyLysGlyGlnAlaIleGluTyrSerGlyLeuThrGlyThrGln 506
                |||                              ||| |||
Db      238 TTTAAA----------------------------TTAGTTGAAGAATAC 258
Qy      507 LeuArgAlaAlaThrGlnLeuAlaIleTyrTyrPheThrAspSerAlaGluLeuAspLys 526
                :::||:::    :::
Db      259 ATTAAAGTACCACTGAAA---------------GACCGAAGAGATATATATT 297
Qy      527 AspLysLeuLysAspTyrHisGlyPheGlyAspMetAsnAspSerThrLeuAlaValAla 546
                :::|||||||||:                   ||||::::
Db      298 GTGTCAATGAAAGATTTATATC-----TTTGAAGAGTTATTTGATAAA----- 345
Qy      547 LysIleLeuValGluTyrAlaGlnAspSerAsnProProGlnLeuThrAspLeuAspPhe 566
                |||:::|     ::::::
Db      346 AAATTAYGTATTGATGAATGTGAAAATGTCAA--------------------- 378
Qy      567 PheIleProAsnAsnAsnLysTyrGlnSerLeuIleGlyThr--------------- 580
                ||||||     ||| |||             :::
```

```
Qy      467 ThrValLysProArgAspThrAspProAspThrPheLeuLysHisIleLysLysValIle 486
                |||:||||||||||:::||:::
Db      199 ACAATCAAGGAAATTGATATTAAACCA---------------AAAAGGGTTGTA 237
Qy      669 GlyLeuProGluGlyTyrSerTyrLeuValLysGluThrAspSerGluGlyTyrLysVal 688
                |||                                      |||
Db      691 GGAGAGTTTGAACATTAT----------------------------GGAAATGTTCAA 720
Qy      689 LysValAsnSerGlnGluValAlaAsn-----AlaThrValSerLysThrGly 704
                ::::||     ||||||                  |||
Db      721 TTCATTAAATTGTCAGAACAGTTCTAACATATTGAATGTTGGATTTGTTAAATCATTATTA 780
Qy      705 IleThrSerAspGluThrLeuAlaPheGluAsnAsnLysGluProValValProThrGly 724
                ||||||||:::                                ||| |||
Db      781 ATTACCAACTGTGAAAGTATTTGATTTTAGTGTCAATAGA------ATTATAGGG 828
Qy      725 ValAspGlnLysIleAsnGlyTyrLeuValAlaLeuIleValIleAlaGlyIleIleSerLeuGly 744
                :::|||       :::|||           ::::||
Db      829 TTATTAAAAAGTGTTGACAGTTAT------------------------GTAGGG 858
Qy      745 IleTrpProGlyIleHisThr 750
                ||||||||||||||||||||
Db      859 TATTGGGGTATTCACACT 876

Db      499 ATTCCAGAAAAACAAAAGAATATTTATAAAGAAGGATAACAGAAATA--------- 540
Qy      600 IleProValThrHisAsnLeuThrLeuArgLysThrValThrGlyLeuAlaGlyLysAspArg 619
                ||||||                      ||| 
Db      541 ------------------------ATAGAAATAAATAATTAACAAAAAAGATATA--- 570
Qy      620 ThrLysAspPheHisPheGluIleGluLeuLysAsnAsnLysGlnGluLeuSerGln 639
                ||||||:::    ||| :::
Qy      640 ThrValLysThrAspLys------------------------------ThrAsnLeuGlu 649
                :::|||||||||||:::                                ::::||:::
Db      571 GAAATAGAACAACAGTAAGAAACATATATTAATAATAGAAATATTCATGACAGTAAATATTAAA 630
Qy      650 PheLys---AspGlyLysAlaThrIleAsnLeuLysHisGlyIleGlySerLeuThrLeuGln 668
                ||||||   ||||||:::     ::::::|||||||||:::
Db      631 ATTAAAAGTGATATTAATAGTCGATGTTACAATAGAAAATGGTGAAGAAAATATGTTTTAT 690
```

```
Search completed: August 19, 2003, 18:20:05
Job time : 3365 secs
```